Association for the Advancement of Animal Breeding and Genetics

Proceedings of the Twenty-first (21st) Conference

Lorne, Victoria, AUSTRALIA
28th September – 30th September 2015

© Association for the Advancement of Animal Breeding and Genetics, 2015
All rights reserved except under the conditions described in the Australian Copyright Act 1968 and subsequent amendments, no part of this publication may be reproduced, stored in a retrieval system or be transmitted in any form, or by any means, electronic, mechanical, photocopying, recording, duplicating, or otherwise, without prior permission of the copyright owner. Contact the Committee of the Association for the Advancement of Animal Breeding and Genetics for all permission requests.

ISSN Number: **1328-3227**
ISBN Number: **978-0-646-94554-5**

Produced by:
Association for the Advancement of Animal Breeding and Genetics
C/- AAABG Secretary
AgriBio | 5 Ring Rd | Bundoora | VIC 3083

Internet web site: http://www.aaabg.org

PRESIDENT'S MESSAGE

A very warm welcome to the AAABG's coming of age meeting, the 21st conference in Lorne, Victoria. Previous Victorian conferences have been held in Melbourne in 1981, 1991 and 2003 but this time we hope you enjoy a chance to visit the Great Ocean road, a famous tourist destination, and enjoy a swim in the pleasantly, cool and refreshing Southern ocean!

The scientific program reports a wide variety of research and I trust contains much of interest to everyone. We have reached the stage where genomic selection is no longer novel and is being imbedded in traditional breeding programs. There is still much to do to get the most out of DNA technology and I hope the conference helps a little to push this topic forward.

Since the first conference in 1979, AAABG has established itself as the premier meeting of Australian and New Zealand scientists working on the genetic improvement of livestock. It has always been intended that AAABG was also an opportunity for scientists and breeders to meet and discuss livestock breeding. However, over the years the number of conferences covering this field has increased and it is not obvious how to structure the conference to meet both objectives. I hope you will come to the OGM to discuss how AAABG should proceed in the future.

I wish to thank the sponsors who have supported the conference, allowing us to invite several overseas speakers and the conference organisers ASN and especially to thank the committee who have organised the Lorne meeting.

Mike Goddard
President, AAABG 2015

ASSOCIATION FOR THE ADVANCEMENT OF ANIMAL BREEDING AND GENETICS 2015 TWENTY FIRST CONFERENCE

COMMITTEE

President	Mike Goddard
President Elect	Toni Reverter
Vice President	Ben Hayes
Secretary	Jennie Pryce
Treasurer	Kath Kemper
Sponsorship coordinator	Hans Daetwyler
Industry Liaison	Matt Shaffer
Editor	Iona MacLeod
Proceedings production	Iona MacLeod
	Bill Pattie
Professional Conference Organiser	ASN Events

CITATION OF PAPERS

Papers in this publication should be cited as appearing in the Proceedings of the Association for the Advancement of Animal Breeding and Genetics (Abbreviation: Proc. Assoc. Advmt. Anim. Breed. Genet.)

For example:

Bowley F.E., Amer P.R. and Meier S. (2013) New approaches to genetic analysis of fertility traits in New Zealand dairy cattle. Proc. *Assoc. Advmt. Anim. Breed. Genet.* **20**: 37-40.

REVIEWERS and SECTION EDITORS

All papers, invited and contributed, were subjected to peer review by two referees.
We acknowledge and thank those people listed below for their work in reviewing the papers (and apologise if we have inadvertently omitted any reviewer from the list). Our special thanks also to the chairs that oversaw the review process as 'section editors' (below).

Section Editors: Kim Bunter, Tim Byrne, Hans Daetwyler, Susanne Hermesch, Kathryn Kemper, James Kijas, David Nation, Wayne Pitchford, Suzanne Rowe, Matt Shaffer, Alison van Eenennaam.

Reviewers:

Mary Abdelsayed
Hassan Aliloo
Hawlader Al-Mamun
Peter Amer
Paul Arthur
Alex Ball
Robert Banks
Stephen Barwick
Tracie Bird-Gardiner
Hugh Blair
Vinzent Boerner
Forbes Brien
Luiz Brito
Daniel Brown
Kim Bunter
Timothy Byrne
Yizhou Chen
Sam Clark
Schalk Cloete
David Cottle
Neil Cullen
Hans Daetwyler
Brian Dalrymple
Kath Donoghue
Cecilia Esquivelzeta-Rabell
Marina Fortes
Mohammad Ferdosi
John Gibson
Michael Goddard
Cedric Gondro
Tom Granleese
Johan Greeff
Mekonnen Haile-Mariam
Bevin Harris
Sue Hatcher
Ben Hayes
Fiona Hely
John Henshall
Susanne Hermesch
Janelle Hocking Edwards
Nicholas Hudson
Howard Jeremy
Patricia Johnson
Arjan Jonker
Matthew Kelly
Kathryn Kemper
Majid Khansefid
James Kijas
Sandra Kittelmann
Lambros (Ross) Koufariotis
Stephen Lee
Andres Legarra
Yutao Li
Li Li
Nicolas Lopez-Villalobos
Duc Lu
Russell Lyons
Iona MacLeod
Keith Macmillan
Christian Maltecca
Karen Marshall
Daniel Martin-Collado
Cecile Massault
John McEwan
Noirin McHugh
Karin Meyer
Stephen Miller
Jasper Munro
Nasir Moghaddar
Suzanne Mortimer
Ellison Musimuko
David Nation
Marina Naval Sanchez
Thuy Nguyen
Miguel Perez-Enciso
Wayne Pitchford
Laercio Porto-Neto
Jennie Pryce
Deirdre Purfield
Ian Purvis
Antonio Reverter
John Roche
Suzanne Rowe
Lise Sandenbergh
Bruno Santos
Helen Smith
Eva Strucken
Bolormaa Sunduimijid
Andrew Swan
Peter Thompson
Bruce Tier
Irene van den Berg
Julius van der Werf
Alison Van Eenennaam
Christy Vander Jagt
Bram Visser
Claire Wade
Sam Walkom
Bradley Walmsley
Tingting Wang
Matt Wolcott
Rob Woolaston

SPONSORS of the 21st AAABG Conference 2015
The financial assistance of the following organisations is gratefully acknowledged.

Major Conference Supporters

Conference Supporters

Session Sponsors

Exhibitors

Others

AAABG was formerly known as the Australian Association for Animal Breeding and Genetics. Following the 1995 OGM the name was changed when it became an organisation with a joint Australian and New Zealand membership. The Association for the Advancement of Animal Breeding and Genetics is incorporated in South Australia.

THE ASSOCIATION FOR THE ADVANCEMENT OF ANIMAL BREEDING AND GENETICS INCORPORATED

OBJECTIVES

(i) to promote scientific research on the genetics of animals;
(ii) to foster the application of genetics in animal production;
(iii) to promote communication among all those interested in the application of genetics to animal production, particularly breeders and their organisations, consultants, extension workers, educators and geneticists.

To meet these objectives, the Association will:

(i) hold regular conferences to provide a forum for:
(a) presentation of papers and in-depth discussions of general and industry-specific topics concerning the application of genetics in commercial animal production;
(b) scientific discussions and presentation of papers on completed research and on proposed research projects;
(ii) publish the proceedings of each Regular Conference and circulate them to all financial members;
(iii) use any such other means as may from time to time be deemed appropriate.

MEMBERSHIP

Any person interested in the application of genetics to animal production may apply for membership of the Association and, at the discretion of the Committee, be admitted to membership as an Ordinary Member.

Any organisations interested in the application of genetics to animal production may apply for membership and, at the discretion of the Committee, be admitted to membership as a Corporate member. Each such Corporate Member shall have the privilege of being represented at any meeting of the Association by one delegate appointed by the Corporate Member.

Benefits to Individual Members

- While it is not possible to produce specific recommendations or “recipes” for breeding plans that are applicable for all herd/flock sizes and management systems, principles for the development of breeding plans can be specified. Discussion of these principles, consideration of particular case studies, and demonstration of breeding programs that are in use will all be of benefit to breeders.
- Geneticists will benefit from the continuing contact with other research workers in refreshing and updating their knowledge.
- The opportunity for contact and discussions between breeders and geneticists in individual members’ programs, and for geneticists in allowing for detailed discussion and appreciation of the practical management factors that often restrict application of optimum breeding programs.

Benefits to Member Organisations

- Many of the benefits to individual breeders will also apply to breeding organisations. In addition, there are benefits to be gained through coordination and integration of their efforts. Recognition of this should follow from understanding of common problems, and would lead to increased effectiveness of action and initiatives.
- Corporate members can use the Association as a forum to float ideas aimed at improving and/or increasing service to their members.

General Benefits

- Membership of the Association may be expected to provide a variety of benefits and, through the members, indirect benefits to all the animal industries.
- All members should benefit through increased recognition of problems, both at the level of research and of application, and increased understanding of current approaches to their solution.
- Well-documented communication of gains to be realised through effective breeding programs will stimulate breeders and breeding organisations, allowing increased effectiveness of application and, consequently, increased efficiency of operation.
- Increased recognition of practical problems and specific areas of major concern to individual industries should lead to increased relevance of applied research.
- All breeders will benefit indirectly because of improved services offered by the organisations which service them.
- The existence of the Association will increase appreciably the amount and use of factual information in public relations in the animal industries.
- Association members will comprise a pool of expertise – at both the applied and research levels – and, as such, individual members and the Association itself must have an impact on administrators at all levels of the animal industries and on Government organisations, leading to wiser decisions on all aspects of livestock improvement, and increased efficiency of animal production.

CONFERENCES

One of the main activities of the Association is the Conference. These Conferences will be structured to provide a forum for discussion of research problems and for breeders to discuss their problems with each other, with extension specialists and with geneticists.

ASSOCIATION FOR THE ADVANCEMENT OF ANIMAL BREEDING AND GENETICS
FELLOWS OF THE ASSOCIATION

"Persons who have rendered eminent service to animal breeding in Australia and/or New Zealand or elsewhere in the world, may be elected to Fellowship of the Association…"

Elected February 1990
R.B.M. Dun
F.H.W. Morley (deceased)
A.L. Rae (deceased)
H.N. Turner (deceased)

Elected September 1992
K. Hammond

Elected July 1995
C.H.S. Dolling
J.R. Hawker
J. Litchfield

Elected February 1997
J.S.F. Barker
R.E. Freer

Elected June 1999
J. Gough
J.W. James

Elected July 2001
J.N. Clarke
A.R. Gilmour
L.R. Piper

Elected September 2005
B.M. Bindon
M.E. Goddard
H.-U. Graser
F.W. Nicholas

Elected September 2007
K.D. Atkins
R.G. Banks
G.H. Davis

Elected September 2009
N. Fogarty
A. Fyfe
J. McEwan
R. Mortimer
R. Ponzoni

Elected September 2011
B.P. Kinghorn
A. McDonald

Elected October 2013
H. Burrow
P. Fennessy
G. Nicoll
P. Parnell

Elected October 2015
P. Arthur
D. Johnson
K. Meyer
B. Tier
R. Woolaston

Citations for the 2015 Fellowships are presented in the following pages.

HONORARY MEMBERS OF THE ASSOCIATION

"Members who have rendered eminent service to the Association may be elected to Honorary Membership…"

Elected September 2009
W.A. Pattie
J. Walkley

PAUL ARTHUR

Originally from Ghana, Paul Arthur completed his postgrad research and worked in Canada before commencing his very prolific career with the New South Wales Department of Primary Industries. Paul is an expert in animal breeding and genetics with a strong national and international reputation, and simultaneously has made an enormous contribution in leadership and direction of animal production research within the NSW public service.

Paul completed his M.Sc. thesis on the use of large dairy breeds in crossbreeding for range beef production, and Ph.D. thesis on the nature, genetics and physiology of double-muscled cattle, at the University of Alberta, Edmonton. The winters of Canada finally proved too long and cold for Paul and he moved his young family to sub-tropical Grafton, NSW, to join the research team at the then NSW Agriculture Grafton Research Station, and became immersed in publication of results from the decade-long beef-cattle crossbreeding and growth research projects being conducted in NSW.

Then a move to Trangie and Paul took on leadership of the Net Feed Conversion Efficiency project, and then onto Sydney where Paul's emerging skills in leadership and management were recognised and he was appointed Director of the Elizabeth Macarthur Agricultural Institute. His research output never slowed and Paul was invited to work in cattle feed efficiency projects in Japan and France, and pigs in Australia. Over the past decade Paul has been a leader in research into the genetics of greenhouse gas emissions in beef cattle, and development of genetic technologies to reduce methane emissions from Australian beef cattle.

Paul has achieved a number of career highlights. He earned promotion to the rank of Senior Principal Research Scientist, the highest rank available to a public-service scientist, with more than 250 scientific publications to his name, including one paper now ranked 3rd among the top 50 most-frequently cited papers ever published by the Journal of Animal Science. He has been awarded the Public Service Medal (PSM), under the Australian Honours System, for "Outstanding public service in the field of animal breeding and genetics". The cattle methane research was runner up for 2013 Eureka Award for Sustainable Agriculture.

To recognise Paul's quiet determination, hard work, leadership and achievement in animal breeding and genetics research, the Association for the Advancement of Animal Genetics and Breeding is pleased to enrol him as a Fellow of the Association.

DAVE JOHNSON

Dave Johnson was born and bred in Southland, New Zealand. After completing his BSc and MSc at Otago University, Dave undertook a PhD at the University of Toronto (his thesis was titled "The symmetric structure theorem for reductive Lie algebras"). He then joined Ag Research as a research statistician: analysing numerous field experiments, as well as providing animal breeding expertise to New Zealand's sheep and beef cattle breeding programs.

Dave also worked in the area of variance component analysis during his years at Ag Research, and this lead to the development of variance component software based on the AI REML algorithm. The software enabled both univariate and multivariate analyses and was subsequently used to estimate genetic and phenotype parameters from the data recorded in progeny test herds at Livestock Improvement. These analyses were based on 100,000s of records collected on multiple traits: analyses that would have been computationally infeasible with any other available software at that time. Many of the genetic and phenotype parameter estimates are still in use in the current national genetic evaluation.

Dave joined Livestock Improvement in 1993 and continued there until now. Over this period Dave has made a significant contribution to New Zealand's dairy cattle genetic evaluation. In the early 1990s many countries, including New Zealand, were adopting the animal model methodology for routine genetic evaluation. During the introduction of the animal model, Dave developed a new methodology to predict total lactation yields from individual test-day information thus providing phenotypic production records for the mixed model analysis. These records accounted for any number of herd tests over any testing frequency and allowed for variable information among herd-mates and for the effects of culling.

Dave was involved in research to improve the methods for estimating the reliability of estimated breeding values. Exact reliabilities can be calculated from the inverse of the mixed model equations. However, in national evaluations, the mixed model contains more than 10 million equations, making them computationally infeasible to invert. A new method of approximating reliability that was computationally fast and provided estimates with low amounts of bias was developed and published by Dave Johnson and Bevin Harris in 1998. This method has been extended to several complex models including test-day and genomic selection models, and to estimation of reliability for Interbull multiple across country sire genetic evaluations (MACE).

In 2007, Dave was an integral member of the team that developed a test-day model (TDM) to provide a national genetic evaluation for dairy production traits. One aspect of the TDM development that has gone unrecognised was Dave's on-demand TDM build for herd-testing customers. The national TDM model is only run approximately every 3 weeks, but farmers required a system to provide updated results at the time of an individual herd-test to enable breeding and culling decisions based on the most up-to-date information. Dave developed a computationally simple solution that incorporated the latest herd-test results into the most recent TDM evaluation, allowing up to 1000 herds (300,000 cows) to be processed daily in peak season.

Over recent years, considerable research effort has been directed towards the application of genomic selection in a national evaluation system for a number of livestock species, including dairy cattle. In New Zealand there was the additional complication of requiring an across-breed genomic evaluation system in order to get genomic evaluations on progeny-tested Jersey Holstein-Friesian crossbred sires. Dave made valuable contributions to a method for the prediction of breeding values incorporating genomic information in an across-breed evaluation: the novel component being the estimation of the genomic relationship matrix in the context of a multi-breed population.

KARIN MEYER

Originally from Germany, Karin completed her training in quantitative genetics with a PhD program at the University of Edinburgh. It would be difficult find three more outstanding people to have had as supervisors in one place than hers of Bill Hill, Alan Robertson and Robin Thompson. Her studies there sparked her lifelong interest in characterising genetic variation.

A series of short-term post-doctoral appointments in Australia, Canada and Edinburgh followed. These appointments generally focussed on specific problems. Indeed, her outstanding ability to assess a task, prepare and analyse the data and, most importantly, complete the study by publication in a scientific journal was recognised widely and her talents were in high demand. Eventually Karin settled into a full time position at AGBU where she still plies her craft.

Estimating variance components for unbalanced data and writing software to do so are the centre pieces of her career. Her work generally involved examining alternative models methods for the analysis of very large sets of data. Generally it required writing the software to complete the analysis as 'off the shelf' programs were unavailable. While at AGBU her primary focus has been the analysis of Beef cattle data, and a seminal paper describing a series of alternative models for analysing data with maternal effects is still cited today. She is, and has been, a key player in the team at AGBU involved with the development of its beef (BREEDPLAN) genetic evaluation system. Her theoretical and practical contributions to characterising genetic variation in livestock have also been appreciated by evolutionary biologists and the plant breeding community where she has made regular contributions over the last 15 years.

Providing software was essentially a 'spinoff' of having solved her own problems but for colleagues, it was often their introduction to Karin and her methods. She was in the vanguard when sire models were replaced by animal models. With DFREML, she played the central part in providing the animal breeding community with the tool they needed. The early 1990s saw new algorithms for maximising likelihoods – some developed by Karin herself – which were quickly incorporated into DFREML. In the late 1990s a new method for analysing longitudinal data – random regression – was added to DFREML. Karin has always been interested in getting the best out of the data. Advances in computing technology has meant that more difficult questions could be posed, but there has always been an underlying goal of having her programs run more quickly. Regular enhancements were made to DFREML and in 2006 Karin released a new incarnation called WOMBAT which she continues to enhance today.

By providing tools for the job, Karin has had an immeasurable impact on the animal breeding community and other related communities. Some indicators of her output include more than 21,000 downloads of WOMBAT since its release, over 90 refereed scientific articles with more than half as sole author, and 51 papers presented to the AAABG including a number of invited papers. Highlights of her career to date were the award of a D.Sc for her contributions to the estimation of variance components by the University of Edinburgh in 2002, her addition to the list of highly cited researchers in 2004 and her elevation to Professor in 2013. Her ability to focus on a task until it is completed with one or more publications is an example to all young scientists.

BRUCE TIER

After science studies at the Australian National University and service with the Department of Foreign Affairs, Bruce Tier completed a Bachelor of Agricultural Sciences at the University of Western Australia in 1980, majoring in Plant Breeding and Agricultural Economics subjects. Raised on a diet of logic puzzles at the family dinner table, Bruce became interested in computer science and programming early in his studies and insisted on including respective courses in his degree schedule, despite discouragement from the faculty. Bruce began his career at AGBU in October 1981, making him the longest serving staff member.

Affectionately known as "Dr. Thong", Bruce was awarded a Ph.D. from the University of New England in 1999 and became a full professor in 2011. In his spare time, Bruce is an avid and competitive Bridge player at national level.

During his 34 years at AGBU, Bruce has played a pivotal role (together with Hans Graser) in the development and implementation of genetic evaluation schemes for Australian livestock. Indeed, Bruce more or less single-handedly devised, built and maintained the software engines driving genetic evaluation for beef (BREEDPLAN) and sheep (OVIS) and provided major inputs to schemes for pigs (PIGBLUP) and trees (TREEPLAN), among others. In this time, BREEDPLAN progressed from a multi-trait analysis of three traits to twenty-five or more traits and, for the larger breeds, millions of animals in the pedigree.

This expansion posed major computational challenges and would not have been feasible without Bruce's analytical and programming skills which allowed him to develop his own, highly efficient strategies – unsurpassed worldwide – to cope with them. Currently, Bruce is implementing the next generation of BREEDPLAN and OVIS, incorporating genomic information through the so-called single-step method, having postponed retirement to do so. No doubt, he will deliver another Rolls-Royce of genetic evaluation schemes to put, yet again, cutting edge methodology for genetic improvement at the fingertips of Australian livestock producers.

Moreover, Bruce has an impressive record of scientific publications in refereed journals and has been highly active in disseminating research results at conferences. He first attended a AAABG meeting in 1984 and has been an author on 78 AAABG conference papers (until 2013), which speaks volumes for his involvement with the society. His key journal papers not directly related to genetic evaluation addressed diverse topics ranging from efficient REML estimation, fast calculation of inbreeding coefficients and gametic imprinting to one of the first studies on the use of multiple genetic markers.

Bruce has provided essential and substantial leadership in scientific research in a range of areas, both within AGBU and with collaborating institutions. For instance, he has been a key person in both the Beef and Dairy Collaborative Research Centres and served on the advisory board of the Australian Dairy Herd Improvement Scheme. Bruce has been instrumental in developing strategies to deal with the avalanche of genomic data in AGBU, and made contributions to problems of genome scans, haplotyping, genotype imputation and polled horn testing, to name a few.

Last but not least, Bruce's impact as a supervisor and colleague has been immeasurable. Known for his irreverence and critical thinking, he has been and is a great team player, willing to work with all sorts and unstinting with his time, advice and, where needed, hands-on assistance to anyone asking, may it be colleagues, students, visiting scientists or emeriti.

For his enormous contributions to animal breeding and quantitative genetics, reflecting brilliance paired with dedication, the Association for the Advancement of Animal Breeding and Genetics is delighted to elect Bruce Tier to a Fellowship of the Association.

ROB WOOLASTON

Raised at Somerton, NSW, Rob grew up on a livestock and grain property, attended Farrar Agricultural High School, and then completed a B.Sc(Agric) degree at UNSW, graduating with first class honours. He then undertook postgraduate training under the supervision of Professor Euan Roberts. After attaining his PhD in 1975, Rob returned to the family farm and spent 7 years finding out about the vagaries of primary production. A desire to return to research saw him then spend 7 very productive years at the Animal Breeding and Research Institute, Katanning, WA, where he began his studies of genetic variation in Merino sheep. Working with Bob Howe and Roger Lewer, Rob's research outputs and delivery to breeders, had a major impact on Merino breeding in WA during this time.

After a stint as Senior Biometrician in the Tree Breeding Section of Queensland Department of Forestry, in 1989 Rob joined CSIRO as Quantitative Geneticist and Project Leader of the Parasite Resistance group at CSIRO Animal Production in Armidale, NSW. During the ensuing 5 years, Rob initiated and conducted many insightful investigations into the genetics of host-parasite interactions in sheep, with particular focus on *Haemoncus contortus* in Merino sheep. During this period he was also involved in breeding research in Fiji and China.

From 1993 Rob's career in CSIRO moved increasingly into research management, and between 1996 and 2001, he was the Manager of the CSIRO Livestock Improvement Program and the Senior Officer-in-Charge of the CSIRO Pastoral Research Laboratory at Armidale. It was during this time that Rob led the successful bid for the establishment of the Australian Sheep Industries Cooperative Research Centre. In 2001 Rob moved to Brisbane, and between this time, and when he left CSIRO in 2004, Rob filled roles as Deputy Chief, and Acting Chief of the Division of Livestock Industries. This was a particularly exciting time within CSIRO as the molecular technologies began to become available to research into livestock breeding. Under Rob's leadership CSIRO made key strategic investments that led to leadership in the international efforts to map key livestock genomes.

Rob's career since 2004 has focussed on provision of consulting services to the major livestock industries. He has performed roles as MLA R&D co-ordinator and technical advisor, and R&D Manager of Pfizer Animal Genetics, and contributed to many other industry boards and advisory committees. In particular Rob has made a major contribution to the Technical Committee of Sheep Genetics, filling the role as Chair between 2005-2008, and again currently.

Rob is the author of over 130 scientific and technical publications, including invited papers at international conferences in Australia, New Zealand, Canada, UK, USA, Uruguay, China, Pakistan and Indonesia. He has co-edited three books on the genetics of host-parasite interactions and animal breeding.

For Rob's very significant contributions to the Australian livestock industries through his research, research management, and contributions to industry boards and advisory bodies, the Association for the Advancement of Animal Breeding and Genetics is please the elect him as a Fellow of the Association.

HELEN NEWTON TURNER MEDAL TRUST

The Helen Newton Turner Medal Trust was established in 1993 following an anonymous donation to the Animal Genetics and Breeding Unit. The Helen Newton Turner Medal is awarded to provide encouragement and inspiration to those engaged in animal genetics. The Medal is named after Dr Helen Newton Turner whose career with CSIRO was dedicated to research into the genetic improvement of sheep for wool production. The Medallist is chosen by Trustees from the ranks of those persons who have made an outstanding contribution to genetic improvement of Australian livestock.

The Helen Newton Turner Medal was first awarded in 1994 to Associate Professor John James and a list of all recipients to date is given below. The recipient of the Medal is invited to deliver an Oration on a topical subject of their choice. The Oration of the 2013 Medal recipient, Prof. Michael Goddard, is reproduced in these proceedings.

Trustees of the Helen Newton Turner Trust are:

- Dr Richard Sheldrake AM (Chairman), representing NSW Department of Primary Industries
- Professor Brian Kinghorn, representing the University of New England
- Mr Scott Dolling, representing the Association for the Advancement of Animal Breeding and Genetics
- Dr Roly Nieper, Representative of the National Farmers Federation
- Dr Robert Banks, Director, Animal Genetics and Breeding Unit

MEDALLISTS

1994 J.W. James
1995 L.R. Piper
1997 J. Litchfield
1998 J.S.F. Barker
1999 C.W. Sandilands
2001 G.A. Carnaby
2003 F.W. Nicholas
2005 K. Hammond
2007 L. Corrigan
2009 R. Hawker
2011 R. Banks
2013 M. Goddard
2015 A. Gilmour

HELEN NEWTON TURNER AO

HELEN NEWTON TURNER MEDALIST ORATION 2013

LIVESTOCK GENETIC IMPROVEMENT IN THE 21ST CENTURY – OPPORTUNITIES AND CHALLENGES

M.E. Goddard
Faculty of Veterinary and Agricultural Science, University of Melbourne, and
Department of Economic Development, Jobs, Transport and Resources, Victoria.

INTRODUCTION

It is a great honour to receive the Helen Newton Turner medal. I knew Helen at the beginning of my career and towards the end of hers. Her knowledge and passion were an inspiration to all of us. In this Helen Newton Turner oration to AAABG I will attempt to take a broad view of the opportunities and challenges facing genetic improvement of livestock, especially cattle and sheep, in the early part of the 21st century.

We are frequently told that the development of an Asian middle class will increase the demand for livestock products especially meat and dairy products. The frustration for producers is that this increased demand does not translate into higher prices. I suspect this is due to competition among suppliers. There are other suppliers of livestock products including the intensive poultry and pig industries and there are plant based substitutes. However, the price of farm inputs does rise. I can only see further rises in the price of feed grain, water, labour and the costs of compliance with animal welfare and environmental regulation. Therefore, to remain profitable I suggest that farm businesses need a 2% per annum increase in economic efficiency, that is, in the ratio of income to costs.

This implies a doubling of efficiency over 35 years. Is that possible? The poultry meat industry has easily achieved this increase, largely through genetic improvement of broilers. Between 1980 and 2010 the Victorian dairy industry increased production per ha 3 fold. The challenge is to do this again. In northern Australian beef production a doubling of production per ha could be achieved if weaning rates were lifted to 80% with reduced cow losses, on farm growth rates averaged 0.7 kg/day and the feedlot finishing phase was shortened. This is a tall order. Perhaps it could be achieved with a cross breeding program using a small, fertile, adapted dam breed and a fast growing sire breed producing calves that graze a feed source of high nutritional quality?

Doubling economic efficiency could be achieved by a combination of increasing the price for products, decreasing cost per ha and increasing production per ha. I suggest it will have to come mostly from increasing production per ha because I cannot see that the prices for farm outputs will increase enough or that costs for farm inputs will decrease enough. Therefore this paper will concentrate on the opportunities to increase production per ha by genetic improvement. While acknowledging that non-genetic improvements will be important they are not the subject of this paper.

Genetic improvement in economic efficiency depends on breeding objectives being aligned with economic efficiency and the use of tools that will deliver genetic progress.

BREEDING OBJECTIVES

The objective should be to increase profit where profit is understood to include all objectives including non-monetary ones such as safety, animal welfare and environment. In the past the most common mistake has been to ignore some traits such as fertility, health and feed conversion efficiency. Especially if there are unfavourable correlations between these traits and other selection criteria, this risks undesirable and costly changes in these traits that are ignored. Fortunately most industries have moved to a more comprehensive breeding objective. However, there are aspects of the objective that are still ignored. For instance, where the commercial animal is crossbred but selection is within purebreds or where breeding stock are selected in a different

environment to the one where their offspring will be farmed. As discussed later, genomic selection offers a better opportunity to select for the traits in the breeding objective than we have had before.

What changes in breeding objectives can we anticipate in the future? The risk of predicting the future is high and, fortunately, objectives change relatively slowly so selection for today's objectives is satisfactory. However, possible changes might be an increased community concern for animal welfare and environmental outcomes, changes in production systems such as robotic milking, sexed semen, cows with twins, once-bred heifers, adaptation to a hotter , drier climate, inclusion of the effect of one animal on the performance of herd mates (so-called social breeding values), and the use of lines specialised for a certain product or environment. We need to include changes in objectives as soon as they become apparent.

TOOLS FOR GENETIC IMPROVEMENT

The tools considered are choice of breed, crossbreeding, avoidance of inbreeding, selection, transgenesis, use of major genes, specialised lines to account for GxE or differences in objectives and mate allocation.

Breed differences can be large but the choice of the best breeds is a once only improvement. Often obvious changes are made quickly by industry and after that the gain from changing breed is small because each breed has some advantages and disadvantages.

Heterosis is also a once only improvement but is an almost cost free improvement in efficiency that is under-utilised.

If heterosis is nearly always an economic advantage, inbreeding depression is nearly always an economic loss. Inbreeding is an inevitable result of small effective population size (N_e) and many breeds have N_e about 100-200. Consequently, inbreeding slowly increases with the expected outcomes such as a rise in frequency of a succession of recessive abnormalities. The management of these recessives is an increasing problem which needs to be put in the usual economic index approach. That is, selection should be based on estimated breeding value for profit including the effect of recessive abnormalities. This will result in culling of animals carrying the more common recessives and culling animals that carry multiple, undesirable genes. However, overall inbreeding levels should be held down by optimising selection decisions to maximise breeding value while minimizing future inbreeding.

Selection is a major opportunity to drive long term improvement in livestock but it is not a new opportunity and in the absence of new technology we can perhaps not expect sudden improvements in its use. The new technology is genomic selection. This is already adopted in the dairy industry and being introduced in other industries. By itself it is beneficial but it is synergistic with 3 other technologies. Traditionally, recording of performance had to be done on selection candidates or their close relatives. This was a problem where the trait was difficult or inconvenient to measure on stud animals. For instance, meat quality is hard to measure on live animals; disease traits are difficult to select for because stud animals are managed to minimise disease; crossbred performance cannot be recorded on purebred animals; feed intake is too expensive to measure routinely. Genomic selection can potentially overcome these problems because the training population can be separate from the elite breeding population. Therefore, the training population can consist, in part, of commercial animals not closely related to the stud animals. This opportunity may be difficult to utilise because traditionally the cost of recording is paid by the owners of elite or stud animals. However, the opportunity is important enough that we should find a new method of paying for the costs. For instance, we could train a genomic prediction equation using crossbred lambs slaughtered and evaluated for FCE and meat quality and yield. Automatic measurement technology would be a synergistic technology with genomic selection because it could reduce the cost of collecting the data for the training population. Another synergistic technology is reproductive technology that allows reproduction at an early age so that generation interval can be reduced. Since DNA can be obtained from an animal at birth or before, selection

decisions can be made earlier in life when using genomic selection than when selection is based on phenotype.

The technology to make transgenic animals has recently improved with the invention of CRISPR and talens. However, this improvement does not overcome the main obstacles to use of transgenesis in agricultural livestock and I suspect we are still years away from adoption of this technology in livestock.

Genotype by environment interactions (GxE) are not uncommon if we interpret E to include the market for which the livestock or their products are intended, the management system and the physical environment. Therefore, one might expect that efficiencies can be gained by breeding a line of animals for a particular environment – management- market combination (EMM). For instance, a line of dairy cows for cheese production or a line of meat sheep carrying the booroola gene for crossing with Merinos. However, the economies of scale work against this idea. It may be more economical to breed a general purpose line that can be sold to many customers rather than a specialist line that is sold to a few customers. This conflict deserves further consideration.

Most of the traits in the breeding objective of sheep and cattle are quantitative traits controlled by a very large number of genes, most of which have a small effect. However, in a few cases, there are known genes of large effect such as booroola for litter size in sheep and myostatin mutations causing double muscling in cattle and sheep. Few of these genes of large effect are deliberately used partially because they have unfavourable side effects. Generally their logical use would be in a line used for crossbreeding for a particular EMM and so utilisation of them depends on specialised lines as discussed above.

CHALLENGES FOR YOUNG SCIENTISTS

The biggest opportunities to increase the rate of genetic gain are in utilising the synergy between genomic selection, reproductive technology and automatic phenotyping. Therefore we should aim to make genomic selection very accurate by using a large, across breed training population including commercial animals under commercial conditions and automatically recorded for traits in the breeding objective. To achieve high accuracy regardless of breed, we need a Bayesian statistical method rather than BLUP, based on genome sequence data from which we have identified the causal mutations or markers in near complete linkage disequilibria with them. To identify these causal mutations we will need to make use of biological information from which we can predict which sites in the DNA cause an effect on phenotype when mutated. Capturing phenotypic information on commercial animals is partly a technology problem (to make measurement very cheap) but also an organisational problem requiring support from industry leaders. To gain full benefit from this technology we need much cheaper reproductive technology such as JIVET and very cheap DNA testing. Cheap DNA testing will lead to more animals being tested and hence, potentially, a large training population provided phenotypic information on these animals can be captured and used for improving the accuracy of the genomic prediction equation. Among many advantages, this will allow animals to be allocated to their most profitable EMM based, in part, on DNA tests.

CONCLUSIONS

For sheep and cattle farming to remain profitable, the economic efficiency will need to double over the next 35 years mainly by increasing production per ha. Genetic improvement can contribute to this objective by careful choice of breeding objectives, use of new technology such as genomic selection, reproductive technology and automated phenotyping

2015 HELEN NEWTON TURNER MEDALLIST CITATION

DR ARTHUR GILMOUR

Dr Arthur Gilmour has made an outstanding contribution to the genetic improvement of Australian livestock, in particular through his development of and support for, software for analysing complex data for research and implementation.

Arthur joined NSW Agriculture in 1970, and spent 10 years as a Biometrician before completing a PhD at Massey under Professors Al Rae and Robert Anderson. He returned to NSW Agriculture where he continued working to 2009, retiring from the role of Principal Research Scientist with 41 years' service.

In his roles in NSW Agriculture, he assisted countless researchers in the design and analysis of experiments, particularly in sheep and plant breeding, as well as developing software used widely in Australia and overseas. These software tools have become "tools of the trade" for researchers and practitioners. An example of implementation software developed by Arthur is BVEST, which was the genetic analysis tool for LAMBPLAN in its formative decade, configured both for use by LAMBPLAN scanning operators and later for centralised analysis of increasingly large across-flock datasets.

Arthur is perhaps best known for the ASREML software, the development of which was stimulated by Arthur's interaction with Dr Robin Thompson, to apply REML methods efficiently, and with Dr Brian Cullis, leading to greatly enhanced models for analysis of plant breeding data. ASREML grew out of REG, which was attractive to users because it was comprehensive and allowed a wide choice of models – features central to ASREML. ASREML is cited in thousands of publications world-wide, and continues to be developed under Arthur's guidance and with his inputs.

Key to Arthur's contribution has been that the tools are backed by seemingly inexhaustible willingness to help others – in his own words: "My role has been to make new ideas in the area of mixed models accessible to general researchers so they can effectively explore their data."

This willingness to share was encapsulated in the citation for Arthur's 2001 Fellowship of the AAABG: "*Not only has ASREML been made readily available to researchers throughout the world, but a discussion group has also been set up that is better described as 'ask Arthur a question'. His generosity in time to individually answer and his resistance to describing perhaps 50% of the questions as stupid are exemplary.*"

Without accurate estimates of genetic parameters, understanding of how traits work genetically and hence genetic evaluation and improvement, are impossible. Arthur Gilmour's life work has enabled this fundamental task of genetics research, development and implementation to be conducted rigorously and accordingly is an extremely important underpinning contribution to the genetic improvement of livestock (and plants) in Australia.

Footnote:
The full text of Dr Gilmore's oration will be published in the 22nd Proceedings of AAABG 2017.

TABLE OF CONTENTS

Breeding Objectives and Economics 1

Selection strategies for breeding objectives in growing pigs 1
S. Hermesch, M.M.A. Arnal, V. Börner and S. Dominik
A genetic line of *Nile Tilapia* undergoing seven generations of selection under moderate salinity water environment 5
N.P. Thoa, N.H. Ninh, W. Knibb and N.H. Nguyen
Non-linear calving difficulty weightings in the Irish dairy industry 9
F.S. Hely, R. Evans, P.R. Amer and A. Cromie
Extended cow liveweight modelling for beef cattle breeding objectives 13
B.J. Walmsley, M.L. Wolcott, W.S. Pitchford, D.J. Johnston1 and S.A. Barwick
Evaluation of alternative selection indexes for non-linear profit traits approaching their economic optimum 17
D. Martin-Collado, B. Visser, T. Byrne, and P. Amer
A review of the national breeding objective and selection Indexes for the Australian dairy industry 21
T. Byrne, D. Martin-Collado, B. Santos, P. Amer, J. Pryce, and M. Axford
Feed saved breeding values for Australian dairy cattle 25
J.E. Pryce, O. Gonzalez-Recio, G. Nieuwhof, W.J. Wales, M.P. Coffey, B.J. Hayes, and M.E. Goddard
Breeding to improve meat eating quality in terminal sire sheep breeds 29
A. Swan, T. Pleasants , and D. Pethick
Accounting for the cost of reproductive technologies during selection in sheep breeding programs 33
T. Granleese, S.A. Clark, A.A. Swan, J.H.J. van der Werf
Genetic expressions and economic impact of performance Recording in multi-tiered sheep breeding schemes 37
B.F.S. Santos, T.J. Byrne, B. Visser, J.H.J. van der Werf, J.P. Gibson and P.R. Amer

Detecting causal variants 41

Why can we impute some rare sequence variants and not others? 41
B.J. Hayes, P.J. Bowman, H.D. Daetwyler and M.E. Goddard
From sheep SNP chips, genome sequences and transcriptones via mechanisms to improved sheep breeding and management 45
B.P. Dalrymple, V.H.Oddy, J.C. McEwan, J.W. Kijas, R. Xiang, J. Bond, N. Cockett, K. Worley, T. Smith and P.E. Vercoe
Pleiotropic multi-trait genome-wide association reveals putative candidate genes for fatty acid composition in Australian sheep 49
S. Bolormaa, J.H.J. van der Werf , B.J. Hayes, M.E. Goddard, and H.D. Daetwyler
Fine-mapping single nucleotide polymorphisms on *bos taurus* chromosome 26 affecting adipose myristic acid Arthur R Gilmour 53
D. Lu, S.M. Hickey, N.G. Cullen, S. Miller
Effects of tex11 and AR polymorhisms on reproduction and Growth traits in Australian beef cattle 57
*L.T. Nguyen, G.M.F.D Camargo, R.E. Lyons, S.A. Lehnert and M.R.S. F***ortes**

Using protein QTL to disentangle variants effecting protein percentage in milk near the casein complex 61
K.E. Kemper, M.J. Carrick and M.E. Goddard
A comparison of genotypes generated by infinium beadchips and a targeted genotype by sequencing method in cattle 65
A.J. Chamberlain, P. Beatson, B.A. Mason, C. Prowse-Wilkins, C.M. Reich, C. Schrooten, B.J. Hayes, J. Tibbits, and M. Hayden
Identification of loci associated with parasite resistance in Australian sheep 69
M. Al Kalaldeh, J.H.J. van der Werf and C. Gondro
Non-synonimous polymorphism in *HELB* is associated with male and female reproductive traits in cattle 73
M.R.S. Fortes, F.B. Almughlliq, L.T. Nguyen, L.R. Porto Neto, S.A. Lehner
Detecting variants associated with complex traits through changing gene expression in cattle 77
M. Khansefid, J.E. Pryce, S. Bolormaa, Y. Chen and M.E. Goddard
Multi-breed GWAS and meta-analysis using whole-genome sequences of five dairy cattle breeds 81
I.van den Berg, D. Boichard and M.S. Lund

Feed efficiency, methane emission and fertility 85

Visualisation of rumen microbiome 85
G.R. Wood, S. Kittelmann, C. Pinares-Patiño, K.G.Dodds, M.R. Kirk, S. Ganesh, S.M. Hickey, P.H. Janssen, J.C. McEwan and S.J. Rowe
Prediction of residual feed intake from genome and metagenome profiles in first lactation Holstein-Friesian dairy cattle 89
M. Wang, J. E. Pryce, Keith Savin and B. J. Hayes
Associations of rumen volatile fatty acids with phenotypic and genetic variation in methane production traits in Angus cattle 93
H.C. Smith, R.M. Herd, K.A. Donoghue, T. Bird-Gardiner, P.F. Arthur and R.F. Hegarty
Liver *microrna* expression revealed the difference in fat Metabolism from cattle divergently selected for residual feed intake 97
Y. Chen, W. AL-Husseini, I. Barchia, C. Gondro, R.M. Herd, and P.F. Arthur
Genetic selection for litter size in cattle 101
L. R. Piper, B. M. Bindon, A. A. Swan, and H. G. Brewer
Towards a data set to investigate feed efficiency in New Zealand maternal sheep 106
P.L. Johnson, S.P. Miller and K. Knowler
Repeatabilities for methane emission in merino ewes on pasture across different ages 110
S. Dominik and V.H. Oddy
Genetic parameters for methane production and relationships with production traits in Australian beef cattle 114
K.A. Donoghue, T.L. Bird-Gardiner, P.F. Arthur, R.M. Herd and R.F. Hegarty
Genomic estimated breeding values for methane production in Australian beef cattle 118
B.J. Hayes, K.A. Donoghue, C. Reich, B. Mason, R.M. Herd and P.F. Arthur
Divergent selection for methane yield in beef cattle 122
T. Bird-Gardiner, K.A. Donoghue, P.F. Arthur, R.M. Herd, and R.F. Hegarty

Genomic prediction in practice 126

Big is beautiful: biology informed sequence exploitation 126
M. Pérez-Enciso, M. Naval-Sánchez, J. Leno-Colorado and A. Reverter
High throughput genotyping-by-sequencing in livestock - ION AMPLISEQ™ 133
Kristian Ridley, Jeremy Walker, Matthew Hickenbotham
Using genomic information to estimate genotype by environment interactions 137
S.A. Clark , S.H. Lee and J.H.J. van der Werf
Lamb eating quality can be managed by using terminal sires with desirable breeding values for intramuscular fat and shear force 141
J.E. Hocking Edwards, B. Hancock, S. Gill, R. Apps and A. Ball
Genomic best linear unbiased prediction using differential evolution 145
H.A. Al-Mamun, P. Kwan, S. Clark, S.H. Lee, H.K. Lee, K.D. Song, S.H. Lee and C. Gondro
Improving the accuracy of across breed genomic predictions 149
M.E. Goddard, I.M. MacLeod, S. Bolormaa, B.J. Hayes, and K.E. Kemper
Data compression: a new way to infer genomic relationship matrices and highlight regions of interest in commercial lines of broiler chicken 153
N.J. Hudson, R. Hawken R. Sapp and A. Reverter
Relationships between change in trait definition and accuracy of genomic breeding value of type traits in Australian Holstein 157
M. Haile-Mariam, J.E. Pryce and B.J. Hayes
Predicting genomic selection accuracy from heterogeneous sources 161
J.H.J. van der Werf, S.A. Clark and S.H. Lee
Accuracy of genomic prediction for merino wool traits using high-density marker genotypes 165
N. Moghaddar, A.A. Swan and J.H.J. van der Werf

Genetics on the ground 169

Considerations in the use of commercial data for genetic evaluation in beef and sheep in Australia 169
R. G. Banks and B. Tier
What causes the negative genetic correlation between maternal and direct effects for lamb birth weight? 173
H.T. Blair
Genetic correlations between purebred and crossbred performance of Poll Dorset sheep 177
D.J. Brown and J.H.J. van der Werf
Using hump height to estimate *Bos indicus* content in the MSA beef grading model which predicts eating quality 181
M. Kelly, J.M. Thompson, L.R. Porto-Neto, R.E. Lyons and S. S. Moore
Are high Australian profit ranking sires best in all herds? Findings from the feeding the genes project 185
J.M. Morton, R.R. Woolaston, P. Brightling, S. Little, K.L. Macmillan, J.E. Pryce and G.J. Nieuwhof
Selection for reduced mature cow weight will reduce body condition unless accompanied by selection for increased fat 189
W.S. Pitchford, S.J. Lee, K.A. Donoghue and B.J. Walmsley

Estimates of genetic parameters for seasonal weight changes of beef cows 193
Karin Meyer and Ian G. Colditz
Effects of productions level on the expression of breed and heterosis for lactation yields of milk, fat and protein in cows milked once and twice daily 197
F. Lembeye, N. Lopez-Villalobos, J.L. Burke and S.R. Davis
Revisiting total weaning weight as a selection criterion 201
K.L. Bunter and D.J. Brown
Is fibre comfort factor required in Merino breeding programs? 205
S. Hatcher and D.J. Brown
Genetic parameters for eating quality traits of Australian lamb 209
S.I. Mortimer, A.A. Swan, L. Pannier, A.J. Ball, R.H. Jacob, J.H.J. van der Werf and D.W. Pethick

Industry focus **213**

A public: private collaboration to evaluate the potential value of genomic information to a vertically-integrated commercial beef cattle enterprise 213
B.T Welly, M.D. MacNeil, A.R. McClain, R.C. Raymond, E.B. Simpson, and A.L. Van Eenennaam
Production performance of merino and Dohne Merino and lambs in pure or crossbreeding systems 217
S.W.P. Cloete and J.J.E. Cloete
Has the beef genetic improvement pipeline been effective? 221
P.F Parnell
Detailing a beef genetics extension strategy 225
S.J. Lee and W.S. Pitchford
Evaluating dairy herd genetic progress 229
M.M. Axford, P.W. Williams, D.P. Abernethy, and G.J. Nieuwhof
Realising genetic improvement for the extensive livestock Industries as a whole 233
R. G. Banks
Inheritance of tail length in merino sheep 237
J.C.Greeff, L.J.E.Karlsson and A.C.Schlink
Testing an approach to account for daughter misidentification In the estimation of breeding values 241
N. Lopez-Villalobos, J.R. Bryant, B.L Harris and D.J. Garrick
Merino breeding objectives under climate change 245
P. Graham, A.A. Swan and R.G. Banks
SNP-based parentage in an Australian cattle industries context: does one size fit all? 249
R.E. Lyons, S. Buttsworth, D. Waine and M. Kelly
Impact of scanning lean cattle on the genetic correlation between scan and carcase intramuscular fat in Angus and Hereford cattle 253
S.F.Walkom, M.G.Jeyaruban and D.J.Johnston

Plenary Presentations **257**

Integration of genomic information into national cattle and sheep evaluations – past, present and future 257
D.P. Berry, J.F. Kearney, N. McHugh, E. Wall, D. Purfield, and A.R. Cromie

Sole haemorrhage, dermatitis and co. - how genomic information and precise phenotypes help to unscramble Genetic background of health traits in dairy cattle 261
K. Schöpke
The opportunities and challenges of integrating genomics in a broiler breeding program 266
R.J. Hawken, R.L. Sapp, R. Okimoto, J. Chen, R. Borg, C-H. Huang, T. Wing, C. Wang, J. Henshall, and M. Abrahamsen
Genomic prediction using sequence data in a multibreed context 270
M.S. Lund, I. van den Berg and D. Boichard

Poster Presentations 274

Selection patterns for Holstein sires in production-recorded Holstein herds with differing feeding systems 274
K.L. Macmillan, J.M. Morton and J.E. Pryce
Mildly penalized maximum likehlihood estimation of genetic covariances matrices without tuning 278
Karin Meyer
Efficiency of a tactical phenotyping strategy for multi-stage selection 282
C Massault and J.H.J van der Werf
A commercial comparison of ewe breeds for reproduction, wool and lamb growth 286
K. P. Ransom, F. D. Brien and W.S. Pitchford
Reproductive performance of Holstein and Jersey heifers and cows in a pasture-based system in South Africa 290
C.J.C. Muller, J.P. Potgieter, S.W.P Cloete and J.A. Botha
Genetic diversity and population structure of four South African sheep breeds 294
L. Sandenbergh, S.W.P. Cloete, R. Roodt-Wilding, M.A. Snyman and A.E. Van der Merwe
Sheep phylogeography and domestication as inferred f rom complete genome sequences 298
M. Naval-Sanchez, S. McWilliam, A. Reverter, M. Perez-Enciso, N. J. Hudson, J. Kijas
Assessing imputation accuracy using a 15k low density panel in a multi-breed New Zealand sheep population 302
R.V. Ventura M. Lee, S.P. Miller, S.M. Clarke and J.C. McEwan
Estimating the genetic (co)variance explained per chromosome for two growth traits using a half sib data structure in sheep 306
C. Esquivelzeta-Rabell, N. Moghaddar, S. Clark and J.H.J. van der Werf
Runs of homozygosity in swakara pelt producing sheep: Implications on sub-vital performance 310
F.C. Muchadeyi, M.T. Malesa, P. Soma and E.F. Dzomba
Genetic diversity and effective population size of eight Iranian cattle breeds 314
E.M. Strucken, K. Karimi, A.K. Esmailizadeh, N. Moghaddar and C. Gondro
Arthrospira platensis: a novel feed supplement influences gene expression in the heart, kidney and liver of prime lambs 318
A.E.O. Malau-Aduli and A. Kashani
Environmental effects on post weaning fleece traits of a Merino sire evaluation flock 322
S.I. Mortimer and A.E. Casey

Methane emissions estimated based on milk fatty acids of jersey and Fleckvieh x Jersey cows in a pasture-based system 326
S. Abel, C.J.C. Muller and B. Sasanti
Comparison of survival phenotypes in specific dairy production systems 330
B. Visser, K. Stachowicz, J.R. Bryant and P.R. Amer
Modelling of longitudinal liveweight data using regression with legendre and eigenvector functions 334
M. D. Price, M. E. Olayemi and J. B. Bryant
Fibre diameter measured in the post-weaning age window is genetically the same trait as yearling fibre diameter 338
Andrew Swan
Measurement of methane traits in the beef information nucleus cattle 342
P.F. Arthur, R.M. Herd, C. Weber, H. Smith and R.F. Hegarty

Reproduction, behaviour and health 346

Regions impacting inbreeding depression and their association with additive genetic effects for jersey cattle from the United States of America and Australia 346
J. T. Howard, J. E. Pryce, M. Haile-Mariam, C. Maltecca
Genetic and phenotypic relationships between kid survival and birth weight in Australian meat goats 350
M.N. Aldridge, D.J. Brown, W.S. Pitchford
Genetic relationships between breech cover, wrinkle and lamb survival in merino sheep 354
S. Hatcher, D.J. Brown, F.D. Brien and M.L. Hebart
A genome-wide association study of non-additive effects for milk yield and fertility in Holstein and Jersey dairy cattle 358
H. Aliloo, J.E. Pryce, O. González-Recio, B.G. Cocks, B.J. Hayes
Effects of selection for fertility on lactation curves 362
E.M. Strucken , Y.C.S.M. Laurenson and G.A. Brockmann
Benchmarking cow health status with dairy herd summary data 366
K.L. Parker Gaddis, J.B. Cole, J.S. Clay, and C. Maltecca
Proposed changes in the genetic evaluation of dairy fertility in New Zealand 370
K. Stachowicz, G.M. Jenkins, P.R. Amer, J.R. Bryant and S. Meier
Using sequence data to improve accuracy of genomic prediction and QTL discovery for dairy cow fertility 374
I.M. MacLeod, B.J. Hayes, M. Haile-Mariam, P. Bowman, A.J.Chamberlain, C.J Vander Jagt, C. Schrooten and M.E. Goddard
Genomic breeding values of heat tolerance in Australian dairy cattle 378
T.T.T. Nguyen, P.J. Bowman, M. Haile-Mariam, J.E. Pryce, B.J. Hayes
Improving the accuracy of genomic selection for lactation anoestrous interval in Brahman cattle 382
M.L. Wolcott, Y.D. Zhang and D.J. Johnston
Australian sheep breeding values for worm egg count retain predictive power across flocks in the presence of GxE 386
L. Li, A.A. Swan, D.J. Brown and J.H.J. van der Werf

Speed Talkes 390

Bessie: a program for multivariate linear model BLUP and Bayesian analysis of larger scale genomic data 390
V. Boerner and B. Tier
Collection of data for the genetic improvement of health traits in Australian dairy cattle 393
M. Abdelsayed, P. Douglas and J.E. Pryce
Quantifying between animal variation for male semen traits and female egg traits in an ostrich flock used to develop an artificial insemination protocol 397
S.W.P Cloete, M. Bonato, A.M.J. Smith and I.A. Malecki
Genomic predictions for meat colour traits in New Zealand sheep 401
L.F. Brito, S.P. Miller, M.A. Lee, D. Lu, K.G. Dodds, N.K. Pickering, W.E. Bain, F.S. Schenkel, J.C. McEwan, S.M. Clarke
Genotype by birth or rearing type interaction in Merino sheep 405
A. Dakhlan, N. Moghaddar and J.H.J. van der Werf
Genetic analyses on carcase characters of Australian Wagyu beef cattle 409
Yuandan Zhang, Bruce Tier and Robert Banks
Change in accuracy of estimated breeding values for Viascan lean-meat yield when estimated from half-sib records 413
H.J.B. Craig, N.B. Jopson and M.J. Young
Effect of different data exclusion policies on the heritability of fat depth in maternal, Merino and terminal sheep 417
A.J. McMillan and D.J. Brown
How precise are enteric methane emission phenotypes or breeding values estimated from spot flux measures? 421
D.J. Cottle, J.I. Velazco, D. Mayer, R.S. Hegarty and R.M. Herd
Implications of manipulating the ewe live weight penalty in maternal sheep indexes in New Zealand 425
J.I. Kerslake, P.R. Amer, M.J. Young and T.J. Byrne

Statistical methods and computation 429

Computing for multi-trait single-step genomic evaluation of Australian sheep 429
Karin Meyer, Andrew Swan and Bruce Tier
Using random forests to identify SNP associated with leg defect in broiler chicken: impact of correcting for population structures 433
Y. Li, A. George, R. Hawken, R. Sapp, S. Lehnert, and A. Reverter
Missing heritability of adaptation phenotypes in tropical cattle 437
Laercio R. Porto-Neto, Nicholas J. Hudson, William Barendse, Sigrid A. Lehnert and Antonio Reverter
Imputation accuracy measurement and post-imputation quality in imputed SNP genotypes for dairy cattle 441
K.M. Tiplady and R.G. Sherlock
Inbreeding depression in adaptation phenotypes of tropical beef cattle using genome wide data 445
L.R. Porto-Neto, N.J. Hudson, W. Barendse, M.A.R. de Cara, S.A. Lehnert and A. Reverter

Optbr: computationally efficient genomic predictions and QTL mapping in multi-breed populations 449
Tingting Wang, Yi-Ping Phoebe Chen, Kathryn E. Kemper, Michael E. Goddard and Ben J. Hayes
The impact of fitting incorrect models on the partitioning of genetic variance components for bionomial lamb traits 453
P.R. Amer K. Stachowicz, P. Pierard and G.M. Jenkins
Backfat as an environmental descriptor in defining growth rate of the pig: a G×E analysis 457
S.Z.Y. Guy, S. Hermesch and P.C. Thomson
Which genomic relationship matrix? 461
B. Tier, K. Meyer and M. H. Ferdosi
Detection and validation of structural variation in cattle whole-genome sequence data 465
L. Chen, A.J. Chamberlain, C.M. Reich, H.D. Daetwyler and B.J. Hayes
A comparison of genetic connectedness measures using data from the NZ sheep industry 469
J.B. Holmes, B. Auvray, S.A. Newman, K.G. Dodds and M.A. Lee
Evolving to the best SNP panel for Hanwoo breed proportion estimates 473
C. Esquivelzeta-Rabell, H.A. Al-Mamun, S.H. Lee, H.K. Lee, K.D. Song and C. Gondro

Systems biology **477**

What's next in genomics? functional annotation of animal Genomes 477
C.J. Vander Jagt, A.J. Chamberlain, B.J. Hayes, L.C. Marett, M.E. Goddard
A map of bovine long non coding RNA across 18 tissues 481
Lambros T. Koufariotis, Yi-Ping Phoebe Chen, Amanda Chamberlian, Christy Vander Jagt, Ben J. Hayes
Potential role of lncrna cyp2c91-protein interactions on immune diseases and obesity 485
P. Suravajhala, L.J.A. Kogelman, G. Mazzoni and H.N. Kadarmideen
Allele specific expression is pervasive in cattle 489
J. Kijas, B. Barendse, C. Bottema, R. Brauning, A. Chamberlain4, S. Clarke, B. Dalrymple, R. Tellam, B. Hayes, J. McEwan, S. Moore
Connecting gene expression and phenotype – preliminary results from RNA sequencing of 150 lambs 493
S. Bolormaa, R. Behrendt, M.I. Knight, B.A. Mason, C.P. Prowse-Wilkins, L. Slocombe, C.J. Vander Jagt, B.J. Hayes, A.J. Chamberlain and H.D. Daetwyler

Proc. Assoc. Advmt. Breed. Genet. 21:1-4

SELECTION STRATEGIES FOR BREEDING OBJECTIVES IN GROWING PIGS

S. Hermesch[1], M.M.A. Arnal[1,2], V. Börner[1] and S. Dominik[3]

[1] Animal Genetics and Breeding Unit (AGBU), a joint venture of NSW Department of Primary Industries and University of New England, Armidale, NSW, Australia
[2] Agrocampus Ouest, Rennes, France
[3] CSIRO Agriculture Armidale, NSW, Australia

SUMMARY

Breeding objectives (BO) have been extended recently for sire lines in Australia to include average daily gain (ADG), backfat, post-weaning survival (PWS), loin weight and belly weight and some consideration of feed cost as either feed conversion ratio (FCR) or daily feed intake (DFI). This study evaluated six selection strategies for two BO that included either FCR or DFI using genetic parameters previously estimated for Australian populations. Response was expressed for one round of selection and a selection intensity of one which is similar to the annual genetic gain that can be achieved in practice. The predicted response in the BO with FCR varied from $3.61 to $4.59 per pig and from $3.48 to $4.00 for the BO with DFI. The lower response in the BO with DFI was partly due to unfavourable genetic association of DFI with ADG. Although PWS was the most important trait in the BO relative to the genetic variation, response in PWS was less than 0.0009 (or 0.09%) due to limited information available for selection candidates. No genetic associations between PWS and other traits were modelled because this information is currently lacking and response in PWS will depend on its genetic associations with other traits. Adding FCR records to the selection index increased response in the BO by 3.6% only because FCR has multiple favourable genetic associations with other traits. In contrast, selection response in DFI was achieved only when juvenile IGF1 and/or DFI were recorded. Therefore, considering feed costs in the BO with DFI is most effective if DFI is also recorded.

INTRODUCTION

Breeding objectives (BO) have been extended over time to better reflect the economic importance of a wider range of traits. Barwick *et al.* (2011) summarised the development of BO used in beef, sheep and pigs in Australia in their review. In pigs, bio-economic models have been used to define BO which may have hindered extension of BO due to the complexity of the underlying models. Recently, Amer *et al.* (2014) and Hermesch *et al.* (2014) presented an alternative approach to derive the economic value of individual traits directly using independent sub models which facilitates future extensions of BO. Hermesch *et al.* (2014) presented economic values for traits of growing pigs which can be used to setup a BO for sire lines. The relative economic importance of traits was outlined based on the genetic standard deviation of each trait indicating the importance of post-weaning survival (PWS) for selection decisions. However, predicted response from different selection strategies was not evaluated by Hermesch *et al.* (2014). The aim of this study was to compare six selection strategies for two BO that are relevant for Australian sire lines.

MATERIALS AND METHODS

The BO included average daily gain (ADG), backfat (BF) and feed conversion ratio (FCR) or daily feed intake (DFI). Further, PWS as well as loin and belly weight (LW, BW) were considered. Economic values for the BO traits were based on Hermesch *et al.* (2014) and Hermesch and Jones (2010, Table 1). Two BO were considered including either FCR or DFI to take feed costs into account. The economic value for ADG, shown in $/pig, was 0.09 or 0.16 $ per g/day when either

FCR or DFI was part of the BO, respectively. The economic value for ADG differs for these two BO because FCR accounts for savings in feed costs due to higher growth (Hermesch *et al*., 2014).

Six different indexes were compared for both BO. The base index (index 1) included records for ADG and BF only. The number of selection criteria was extended through stepwise inclusion of piglet birth weight (PBW, index 2), PWS (index 3), LW and BW (index 4), juvenile insulin-like growth factor 1 (IGF1, index 5) and lastly FCR or DFI (index 6). Piglet birth weight and IGF1 were considered as selection criteria because both traits have favourable genetic associations with efficient lean meat growth (Hermesch *et al*., 2001; Bunter *et al*., 2005) and are recorded in young growing pigs. Genetic parameters are outlined in Table 1 based on these previous studies outlined above as well as Hermesch (2008). No information was found about genetic or phenotypic correlations between PWS and other performance traits which consequently were assumed to be zero. Index calculations were performed using the MTIndex program of van der Werf (http://www.personal.une.edu.au/~jvanderw).

It was assumed that ADG, BF and PBW were available for the selection candidate, six full sibs and 30 half sibs. Although PWS is available for all animals, only surviving pigs are selected and no distinction can be made between pigs with high or low liability for survival. For this trait, family selection is more effective because it is a threshold character with low incidence (Falconer and Mackay, 1996). Therefore, it was assumed that information about PWS was only available for the sire because the mean reliability for survival of sires is better known based on information about progeny from multiple litters. The carcase traits LW and BW were available for two full sibs and ten half sibs. For IGF1, information was available for the selection candidate, one full sib and ten half sibs. Feed intake is most expensive to measure and it was assumed that FCR or DFI were only recorded on the selection candidate and five half sibs.

Table 1. Genetic standard deviations (GSD), heritabilities (h^2), economic values (EV) and genetic (below diagonal) or phenotypic (above diagonal) correlations for traits.

	GSD	h^2	$EV_{FCR/DFI}^A$	ADG	BF	FCR	DFI	PWS	LW	BW	IGF1	PBW
ADG	30.000	0.31	0.09/0.16[A]		0.11	-0.20	0.32	0.00	-0.14	0.20	0.09	0.38
BF	1.000	0.33	-1.70	0.02		0.06	0.11	0.00	-0.37	0.11	0.06	-0.14
FCR	0.150	0.12	-27.44/0.00[A]	-0.37	0.10		0.00	0.00	-0.14	0.02	0.15	-0.10
DFI	0.094	0.24	0.00/-36.12[A]	0.50	0.35	0.00		0.00	-0.05	0.05	0.09	0.10
PWS	0.038	0.05	182.88	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00
LW	0.680	0.42	3.60	-0.15	-0.54	-0.40	-0.20	0.00		-0.29	-0.05	0.05
BW	0.390	0.27	1.20	0.16	0.30	0.25	0.20	0.00	-0.51		0.05	-0.05
IGF1	13.070	0.21	0.00	0.06	0.21	0.65	0.41	0.00	-0.20	0.20		0.04
PBW	0.064	0.04	0.00	0.56	-0.43	-0.30	0.20	0.00	0.20	-0.20	-0.33	

Trait abbreviations: ADG: average daily gain (g/day), BF: back fat (mm), FCR: feed conversion ratio (kg feed/ kg gain), DFI: daily feed intake (kg/day), PWS: post-weaning survival (0/1), LW: loin weight (kg), BW: belly weight (kg), IGF1: juvenile insulin-like growth factor-I (ng/ml), PBW: piglet birth weight (kg).

[A] Economic values ($/pig) differ for breeding objectives with either DFI (first value) or FCR (second value)

RESULTS AND DISCUSSION

The response to selection is shown per generation assuming a selection intensity of one. This

response is similar to expected annual genetic gains because the selection intensity achieved in practice is similar to the generation interval of about 1.65 years. For the BO with FCR, the overall response was $3.61 per pig for Index 1. This index, which has traditionally been used in pig industries, leads to favourable responses in FCR and LW due to favourable genetic associations with ADG and BF. Individual PBW has a low heritability and recording PBW (index 2) is of limited value for genetic improvement of efficient lean meat growth in growing pigs.

Post-weaning survival was the most important BO trait in both BO accounting for 38% (FCR) or 35% (DFI) of the selection emphasis relative to genetic standard deviations of traits. Using information about PWS for the sire in index 3 resulted in a predicted response of 0.0009 (or 0.09%) which implies that it would take about 12 generations to improve PWS by one percent. The index calculations in this study assumed no genetic associations between PWS and other traits. Additional analyses demonstrated (results not shown) that response in PWS was lowly negative when unfavourable genetic correlations with a magnitude of 0.2 were assumed with other BO traits. Knap (2014) demonstrated favourable genetic trends for survival of pigs from birth to slaughter based on combined pre- and post-weaning survival. Genetic trends for PWS were not explicitly shown. It is therefore important to estimate genetic associations between PWS and other performance traits to monitor genetic trends in PWS better and to establish whether genetic improvement of PWS is feasible.

Table 2. Traits measured in index, accuracy of index (Acc), overall selection response (ΔG in $/pig) and response in breeding objective traits per generation with selection intensity of one – breeding objective includes feed conversion ratio instead of daily feed intake.

Index	Traits measured[1]	Acc	ΔG	ADG	BF	FCR	PWS	LW	BW
1	ADG, BF	0.361	3.61	15.63	-0.467	-0.036	0.00000	0.121	-0.0228
2	Index 1 + PBW	0.364	3.63	15.90	-0.466	-0.037	0.00000	0.119	-0.0202
3	Index 2 + PWS	0.372	3.72	15.55	-0.455	-0.036	0.00090	0.116	-0.0197
4	Index 3 + LW + BW	0.414	4.13	13.75	-0.429	-0.047	0.00081	0.224	-0.0611
5	Index 4 + IGF1	0.444	4.43	12.99	-0.416	-0.062	0.00075	0.220	-0.0644
6	Index 5 + FCR	0.460	4.59	12.62	-0.396	-0.069	0.00073	0.230	-0.0684

[1] for trait abbreviations see Table 1.

Adding information about LW and BW led to the highest marginal gain in the overall BO with FCR. The response increased by 11.0% from 3.72 to 4.13 $/pig due to genetic gain in LW for the BO with FCR. No favourable response was achieved in BW due to unfavourable genetic correlations with LW or BF. In comparison, adding IGF1 and FCR to the selection index for the BO with FCR increased the overall response to $4.43 and $4.59 per pig equivalent of an increase of 7.3% and 3.6% relative to the preceding index. Therefore, recording FCR does not lead to substantial additional response once other traits with favourable genetic correlations to FCR (IGF1, LW, ADG and BF) have already been considered.

Responses in ADG and FCR contributed most to the overall response of the BO with FCR accounting for 39% and 27% in index 1, and 25% and 41% in index 6, respectively. As more traits were added to the index, responses in BF decreased while responses in the additional carcase trait LW increased. Backfat and LW accounted for 22% to 15% and 12% to 18% of the overall responses in the BO which demonstrates that selection for carcase traits related to lean meat

content continues to provide economic returns.

Including DFI in the BO is an alternative selection strategy to consider feed cost (Table 3). Selection response in DFI was only achieved after juvenile IGF1 or DFI were recorded (index 4 and 5). All other selection strategies did not lead to any response in DFI due to its unfavourable genetic correlation with ADG of 0.50. Consequently, the response in the overall BO with DFI was lower in comparison to the previous BO with FCR ranging from $3.48 to $4.00 per pig for index 1 to 6. The favourable genetic correlation between DFI and BF implied that more response was obtained in BF in comparison to the BO with FCR. Further, the added response in the BO due to recording an additional trait was highest for DFI contrary to recording FCR in the previous BO. Therefore, considering feed costs in the BO with DFI is most effective if DFI is also recorded.

Table 3. Traits measured in index, accuracy of index (Acc), overall selection response (ΔG in $/pig) and response in breeding objective traits per generation with selection intensity of one – breeding objective includes daily feed intake instead of feed conversion ratio.

Index	Traits measured[1]	Acc	ΔG	ADG	BF	DFI	PWS	LW	BW
1	ADG, BF	0.383	3.48	11.84	-0.583	0.000	0.00000	0.179	-0.049
2	Index 1 + PBW	0.386	3.51	12.15	-0.580	0.000	0.00000	0.176	-0.046
3	Index 2 + PWS	0.396	3.60	11.86	-0.566	0.000	0.00093	0.172	-0.044
4	Index 3 + LW + BW	0.406	3.69	11.47	-0.560	0.000	0.00091	0.226	-0.059
5	Index 4 + IGF1	0.416	3.78	11.30	-0.556	-0.003	0.00088	0.228	-0.062
6	Index 5 + DFI	0.440	4.00	10.13	-0.557	-0.015	0.00084	0.221	-0.064

[1]: for trait abbreviations see Table 1.

ACKNOWLEDGEMENTS

This study was supported by Australian Pork CRC under project 2B-104. Mathieu Arnal contributed to this study while staying at AGBU and CSIRO as a visiting MSc student from Agrocampus Ouest in Rennes, France.

REFERENCES

Amer P.R., Ludemann C.I. and Hermesch S. (2014) *J. Anim. Sci*. **92**: 5345.

Barwick S.A., Swan A.A., Hermesch S. and Graser, H.-U. (2014) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **19**: 23.

Bunter K.L., Hermesch S., Luxford B.G., Graser H.-U. and Crump R.E. (2005) *Aust. J. Exp. Agr*. **45**: 783.

Falconer D.S. and MacKay T.F.C. (1996) 'Introduction to Quantitative Genetics' 4th ed. Longman, Harlow.

Hermesch S. (2008) *Animal* **2**: 1178.

Hermesch S., Bunter K.L. and Luxford B.G. (2001) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **14**: 227.

Hermesch S. and Jones R.M. (2010). In '2010 Pig GeneticsWorkshop Notes', pp. 39-46, editors S.Hermesch and K. Dobos, Animal Genetics and Breeding Unit, Armidale.

Hermesch S., Ludemann C.I. and Amer P.R. (2014) *J. Anim. Sci*. **92**: 5358.

Knap P.W. (2014) 10th *World Congress of Genetics Applied to Livestock Production*, paper 007.

Proc. Assoc. Advmt. Breed. Genet. 21:5-8

A GENETIC LINE OF NILE TILAPIA UNDERGOING SEVEN GENERATIONS OF SELECTION UNDER MODERATE SALINITY WATER ENVIRONMENT

N.P. Thoa[1,2], N.H. Ninh[1], W. Knibb[2] and N.H. Nguyen[2]

[1] Research Institute for Aquaculture No.1, Vietnam
[2] University of the Sunshine Coast, Maroochydore DC, QLD 4558, Australia

SUMMARY

To date, commercial strains of Nile tilapia (*Oreochromis niloticus,* L.) are not available to culture in brackish water systems. Our attempts were to develop a genetic line of Nile tilapia that can perform well not only under a moderate salinity water environment, but also have desired eating and marketing characteristics. The genetic line developed from our study showed a significant improvement in growth performance, with an average genetic gain for body weight of about 8% per generation (one year per generation). Selection for high growth did not have any adverse effects on survival or deformity. Under on-farm testing conditions, our selected line outperformed other strains especially in saline water environments. The overall eating acceptability of the improved genetic line was superior to other freshwater tilapia counterparts. It is concluded that the moderate salinity tolerant tilapia line developed from our selective breeding program can be cultured effectively under a diverse array of farming systems.

INTRODUCTION

Tilapia is the second most important commercial freshwater fish, after carp, for freshwater aquaculture around the world. The species are widely cultured in more than 100 countries (El-Sayed, 2006). Among over 100 tilapia species, about three major species are being used for aquaculture, Nile tilapia (*Oreochromis niloticus,* L.) contributes about 71% of the global tilapia production (4,207,900 metric tons with an estimated value of US$ 6,923 million in 2012) (Fitzsimmons, 2013). Despite the significant role of tilapia in the fishery sector, there has been a lack of improved stocks that can produce high quality seed to supply fish producers and farmers. As a consequence, commercial production is not highly economically viable due to low productivity of existing stocks. A number of breeding programs have been initiated to develop improved lines of Nile tilapia (*O. niloticus*) such as the Genetically Improved Farmed Tilapia (GIFT) strain (Nguyen *et al.*, 2010; Hamzah *et al.*, 2014) or GIFT-derived strain called as NOVIT4 (Luan *et al.*, 2008). However, these strains are suitable for a freshwater culture environment only. To date there are no improved lines of Nile tilapia that can perform well in brackish water systems (> 10 parts per thousand; ppt). In this project, we aimed to develop a fast growing strain of Nile tilapia with high survival and good adaptation to a range of aquaculture production systems.

The specific objectives of our study were: (i) to improve growth performance of the genetic line under saline water environments (10 – 20 ppt), (ii) to explore possibilities to select for new traits such as early survival, delayed maturity or disease resistance, (iii) to examine the effect of genotype by environment interaction, (iv) to evaluate physiological stress response and flesh quality characteristics of the improved strain relative to tilapia counterparts currently available for freshwater culture systems, and (v) to understand osmoregulation of salinity tolerance in this improved strain. In this paper, we present the selection response achieved from the genetic improvement program, specifically selection response to growth rate after seven generations in moderately salinity water. In addition, we highlight some of the main findings obtained from this study in relation to the objectives (ii) to (v).

MATERIALS AND METHODS

Origin of the selection population. The selection population was established in 2007 at Research Institute for Aquaculture No.1 (RIA1), Vietnam, from a complete diallel cross involving three different strains: genetically improved farmed tilapia (GIFT), GIFT-derived strain named as NOVIT4 and Taiwanese Nile tilapia. A detailed description of the population is given in Ninh *et al.* (2014) and Thoa *et al.* (2015).

Family production and performance testing. In each generation from 2007 to 2014, between 66 and 108 full- and half-sib families were produced within a six week period. After hatching, each family was kept separately in hapas (net cages of $1m^3$) until the fish reached a fingerling size of 5-10 g for physical tagging. A random sample of 50-100 individuals were then identified using Passive Implant Transponder (PIT). The tagged fish from all families were communally grown out in a saline water pond of 15 – 20 ppt over a period of about 4 months for successive annual generations between 2007 and 2013. Representatives of each family produced in 2014 were also tested in a freshwater pond environment. At harvest, body measurements were made on all individual fish, including body weight, standard length, body depth and body width. In the latest generation (year 2014), fitness (early survival, sexual maturity) and disease related traits (gill condition, deformity) were also recorded.

Genetic evaluation and selection. Best linear unbiased prediction (BLUP) was applied to a multi-trait mixed model and used to estimate breeding values (EBV) for harvest body weight for all individuals in the pedigree. Based on EBV ranking for harvest body weight, a combined within- and between- family selection was applied to select the best (highest EBV) individuals to become parents of subsequent generations. Across generations, the proportion of females and males selected were about 4.5% and 3.5%, respectively. Selection was practised on body weight at harvest. Mating of closely related individuals was avoided and the number of individuals contributing to next generations was constrained to minimise inbreeding. The average inbreeding rate was less than 0.5% per generation.

Estimation of genetic parameters. Restricted maximum likelihood method (REML) was applied to estimate genetic parameters for traits studied. A linear mixed model was used to analyse continuous traits such as body weight or food conversion ratio. The random effects were the additive genetic effects of individual fish and full-sib family groups and the fixed effects were generation, sex, environment and their two-way interactions. Age from birth to harvest within sex and generation was fitted as a linear covariate for body weight traits. For binary characters including survival or sexual maturity, generalised threshold mixed models were applied as described by Thoa *et al.* (2015) and Nguyen *et al.* (2014).

RESULTS AND DISCUSSION

Response to selection. Genetic evaluation of 36,145 individual animals with performance records showed that direct genetic gain for body weight in the selection program was averaging 8% per generation or 0.20 to 2.84 genetic standard deviation units (Figure 1). Correlated changes in survival rate during grow-out were close to zero (Ninh *et al.* 2014). The moderate heritability for body weight ($h^2 = 0.53 \pm 0.12$) also shows that the present population will continue to respond to future selection.

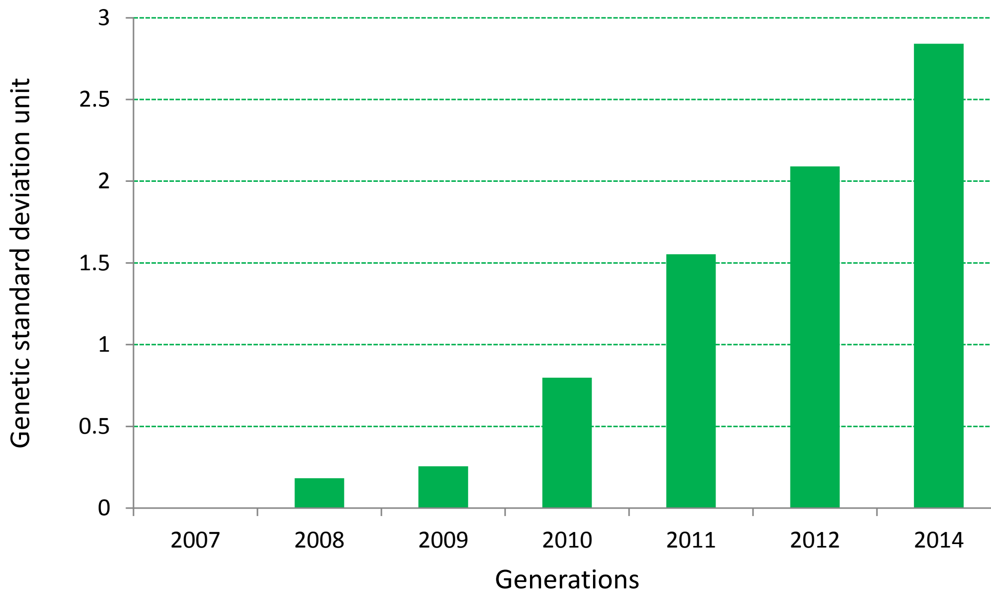


Figure 1. Genetic gain for body weight in standard deviation units achieved in the selection population of Nile tilapia under moderate salinity water environment (10 – 20 ppt) over seven generations including the base population in 2007.

Genetic parameters for new traits. To broaden the breeding objectives for this species, we examined the quantitative genetic basis of new traits collected in the latest generation (2014). REML analysis showed that there are additive genetic components for early survival rate, sexual maturity, gill condition and morphological deformity (heritability range from 0.05 to 0.29), suggesting potential to include these traits in a future breeding program for this population.

Offspring of the salinity tolerance line can perform well under freshwater ponds. We attempted to understand if the moderately salinity tolerant line can perform well in freshwater systems. The data, from performance testing under both salinity and freshwater environments, were combined with a full pedigree including a total of 36,145 individual animals recorded from 2007 to 2014. The estimates of genetic correlations for homologous trait expressions between salinity and freshwater environments were very high and close to one (0.78 – 0.99, s.e. 0.07 to 0.29). This result suggests that the genotype by environment interaction was not significant for growth and fitness related traits in this population. It is also suggested that the moderately salinity tolerant genetic line developed from this selective breeding program can be used in freshwater farming systems.

Physiological response. One indicator of physiological response measured in the selected line was Na^+ and K^+-ATPase activity. Under the same salinity culture conditions, the improved line showed an increase in Na^+ and K^+-ATPase activity relative to other ‘freshwater’ counterparts.

Flesh attributes and eating quality. Preliminary analyses show that in comparison with commercial ‘freshwater’ tilapia strains, our salinity tolerant line had better flesh quality attributes and overall eating quality.

Osmoregulation. Future studies will examine possible changes at the molecular and genomic levels as well as osmoregulatory adaptation that may have been accompanied by the selection program in the present population.

REFERENCES

El-Sayed, A.-F.M. (2006) 'Tilapia culture'. CABI.
Fitzsimmons, K. (2013) World Conference on Tilapia, Rio de Janeiro, Brazil.
Hamzah, A., Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., Yee, H.Y.*, et al.* (2014) *Journal of Tropical Agricultural Science* **37**: 411.
Luan, T.D., Olesen, I., Ødegård, J., Kolstad, K., Dan, N.C. (2008) The proceedings of the 8th International Symposium on Tilapia in Aquaculture, Egypt, pp. 231.
Nguyen, N.H., Ponzoni, R.W., Abu-Bakar, K.R., Hamzah, A., Khaw, H.L.*, et al.* (2010) *Aquaculture* **305**: 1.
Nguyen, N.H., Quinn, J., Powell, D., Elizur, A., Thoa, N.P.*, et al.* (2014) *BMC genetics* **15**: 132.
Ninh, N.H., Thoa, N.P., Knibb, W., Nguyen, N.H. (2014) *Aquaculture* **428-429**: 1.
Thoa, N.P., Knibb, W., Ninh, N.H., Van Dai, N., Nhat, P.H.*, et al. Aquaculture* **442**: 112.

Proc. Assoc. Advmt. Breed. Genet. 21: 9-12

NON-LINEAR CALVING DIFFICULTY WEIGHTINGS IN THE IRISH DAIRY INDUSTRY

F.S. Hely[1], R. Evans[2], P.R. Amer[1] and A. Cromie[2]

[1] AbacusBio Ltd, Dunedin, New Zealand
[2] Irish Cattle Breeding Federation, Cork, Ireland.

SUMMARY

The genetic evaluation of calving difficulty in Ireland is currently based on farmer reported scores. The predicted transmitting abilities (PTAs) can then be converted using non-linear regression on phenotypic calving difficulty for high reliability bulls to an estimate of the percentage of difficult calvings for Dairy Cows and Dairy Heifers. A non-linear penalty is proposed to account for the risk that a low reliability bull may have a much higher percentage of difficult calvings than initially predicted. In addition to this penalty, a non-linear weighting is proposed for calving difficulty within a dairy beef index to reflect that a 1% increase in calving difficulty at a level that is already considered high has a larger economic impact than a 1% increase at a lower level.

INTRODUCTION

Calving difficulty is a trait of economic importance in cattle, with costs including loss of calf, loss of milk for dairy cows and veterinary costs incurred by difficult labour (Dekkers, 1994). Work carried out by Amer *et al.* (2001) determined the economic weighting of calving assistance in the Ireland and incorporated this economic value into breeding objectives for beef. Currently, prototype genetic evaluations for dairy calving difficulty traits are computed by the Irish Cattle Breeding Federation (ICBF) for "Dairy Heifer" (DH) and "Dairy Cow" (DC) matings. While all data (including data from beef matings on dairy cows) is simultaneously used in the multi-trait evaluation, the evaluation produces separate PTAs for heifers and cows. The genetic evaluation system uses calving category scores as input phenotypes in a way that was found to optimise the heritability of the traits and the stability of the evaluation system. However, the resulting PTAs need to be transformed from the underlying 4 point score scale (defined as 1 = no assistance, 2 = some assistance, 3 = considerable difficulty, 4 = veterinary assistance), to a percent difficult scale (defined as percentage of scores 3 and 4), in order to align with economic value calculations and to make more sense to farmers. The translation is undertaken using a non-linear function, separately for each type of calving (e.g. DH versus DC), and is based on comparing PTAs for high reliability bulls with the actual number of difficult calvings in their daughters.

MATERIALS AND METHODS

Two levels of non-linear weightings were integrated into the transformed calving difficulty PTAs, firstly incorporating a non-linear penalty for low reliability bulls and secondly applying a non-linear weighting for percentage of difficult calvings in an index for dairy beef suitability.

Non-linear reliability penalty. The conversion of PTAs from the underlying genetic evaluation scale to the percent difficult scale involves fitting a quadratic to the PTAs for high reliability bulls using their progeny percentage of difficult calvings for DH and DC as the independent variable. However, the direct conversion to a PTA based on percentage of difficult calvings does not take account of the downside risk associated with getting a high percent difficult bull, which is not completely offset by the upside risk of the bull turning out to have lower than expected difficulty. For this reason, an adjustment calculation has been made to translate the raw PTA into a weighted average expected level of percent difficulty.

To do this, each quadratic function was combined with an assumed normal distribution of uncertainty around the PTA. This distribution has a standard deviation combining the breed specific genetic variance (σ^2) calculated from de-regressed PTAs, by breed, and the reliability (ρ^2) of the bull's own PTA. In order to get the expected average percentage of calving difficulty that takes into account the uncertainty of the calving difficulty PTAs, the estimates were weighted by the normal distribution around the point using the integral in the following equation, where x is the PTA.

$$\int_{-\infty}^{\infty} N(x,\mu,\sqrt{\sigma^2(1-\rho^2)})\cdot f(x)\ dx \rightarrow \int_{-\infty}^{\infty} \frac{1}{\sqrt{\sigma^2(1-\rho^2)}\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2(1-\rho^2)}} \cdot (ax^2+bx+c)\ dx$$

This integral simplifies to the following weighting function: $f(x, \sigma^2, \rho^2) = ax^2 + bx + c + a\,\sigma^2(1-\rho^2)$ where the intercept of the quadratic changes depending on the breed specific genetic variance of calving difficulty and reliability of a particular bull of interest. This gives a penalty function of a $\sigma^2(1-\rho^2)$ which can be applied to the transformed PTA. Using this adjustment adds larger weight to PTAs with lower reliabilities in breeds with a high amount of variability, as they have a larger spread, which represents higher risk.

Non-linearity in utility of calving difficulty. A linear weighting on calving difficulty within an index suggests that at any level of calving difficulty, an additional increase of 1% in calving difficulty has the same negative impact for farmers. However, in reality it is more likely that farmers would be more averse to an increase in calving difficulty when the mean level is already high compared to at lower levels. This suggests a non-linear economic weighting transformation would be appropriate for calving difficulty. The justification for this transformation is that with the high rates of assistance, a significant proportion of the herd is compromised in their health and rebreeding success, and a large amount of this cannot be compensated for as easily as a small amount. Barwick *et al.* (2001) investigated employing non-linear selection emphasis on calving ease EBVs in beef and determined that the method was useful in responding to the different levels of calving ease found in beef cattle.

Three non-linear weighting function options for calving difficulty were compared with a traditional linear economic weight derived based on the economic cost involved in the increased stockman hours, veterinary interventions, cow mortality, disposal and infertility as well as loss in milk sales as originally calculated by Berry *et al.* (2005). These economic weights were combined into an index designed to identify profitable beef bulls for use in Irish Dairy herds that combines calving difficulty, calf value incorporating a mortality adjustment and gestation length (McHugh *et al.* 2012).

RESULTS AND DISCUSSION

Non-linear reliability penalty. Figure 1 shows the penalty function as reliability increases from 0 to 1 (0 to 100%) for three different levels of variation which were equivalent to the variances observed in a trial data set of raw Angus (0.05), Charolais (0.09) and Belgian Blue (0.18) bull PTAs from ICBF. When the reliability was low (<0.1), the penalty in the high variability breed was around 3-3.5%, so the percentage of difficult calvings was inflated by the penalty of 3-3.5% to account for the risk that, due to the poor reliability of the PTA, the bull produces more difficult calvings than expected. The risk is lower in the breeds with less variability as the spread in potential PTAs is smaller at the same reliability.

Table 1 shows the genetic variance by breed for the dairy heifer and dairy cow PTAs as derived from the initial set of AI bulls from ICBF. The average penalty applied by breed to these AI bulls is also shown for dairy heifers and dairy cows, along with the average penalty applied to a dataset of young bulls that do not have any progeny records. Table 1 shows that the breeds with

the highest variance in the dairy heifer PTAs were Herefords, Belgian Blue and Simmentals, although of these three breeds only Belgian Blues also had a large variance in the dairy cow PTAs. The variance for the Angus PTAs was reasonably low for both dairy heifer and dairy cow PTAs, and thus the average adjustment was fairly low, with an average of 0.37-0.44% additional in the AI sires, and 0.52-0.63% additional in the young bulls. In comparison, the average adjustment for the Belgian Blue sires was around 1.3% in the AI sires and around 2% extra in the young bulls.

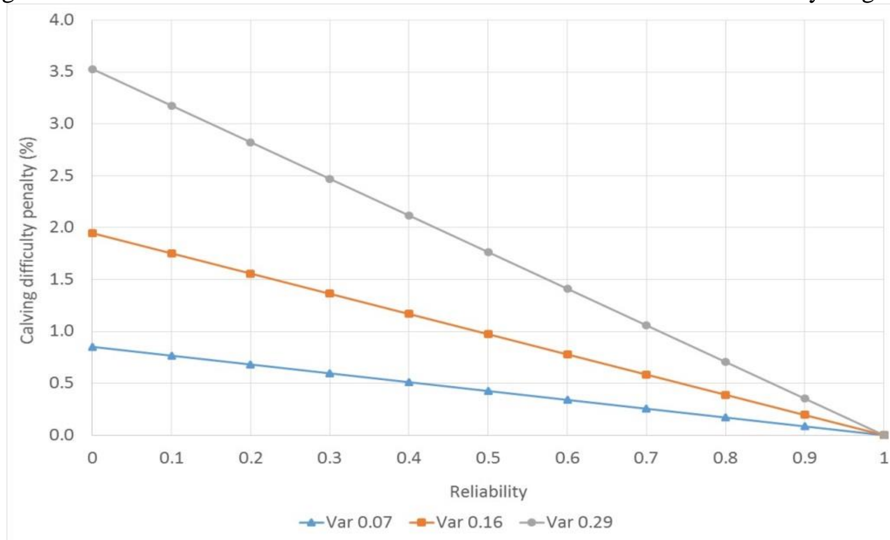


Figure 1. The additional penalty applied in percentage calving difficulty as the PTA reliability increases from 0 to 1 for three different levels of within breed variance.

Table 1. The derived genetic variance, average penalty for an initial set of AI bulls and the average penalty for a set of young bulls with no progeny records, by breed for both dairy heifer and dairy cow evaluations.

Breed[1]	Dairy heifer			Dairy cow		
	Genetic var	Avg sire penalty %	Avg young bull penalty %	Genetic var	Avg sire penalty %	Avg young bull penalty %
AA	0.05	0.37	0.52	0.05	0.44	0.63
AU	0.06	0.54	0.61	0.02	0.17	0.20
BA	0.08	0.60	0.81	0.02	0.19	0.28
BB	0.18	1.37	1.88	0.16	1.30	2.03
CH	0.09	0.69	0.96	0.04	0.29	0.46
HE	0.34	3.40	3.46	0.05	0.58	0.58
LM	0.07	0.40	0.65	0.03	0.15	0.30
PI	0.06	0.58	0.66	0.02	0.17	0.21
PT	0.07	0.61	0.75	0.02	0.19	0.25
SA	0.04	0.29	0.42	0.01	0.10	0.17
SH	0.07	0.66	0.73	0.02	0.28	0.32
SI	0.12	0.90	1.17	0.04	0.36	0.53

[1]Angus (AA), Aubrac (AU), Blonde D'Aquitaine (BA), Belgian Blue (BB), Charolais (CH), Hereford (HE), Limousin (LM) Piedmontese (PI), Parthenaise (PT), Saler (SA), Shorthorn (SH).

Non-linearity in utility of calving difficulty. The three proposed non-linear economic weightings place moderate, strong and very strong non-linear penalties on high calving difficulty bulls (Figure 2). The weighting functions showed a widening difference between the linear and quadratic indexes as the percentage difficulty increases, with bulls whose calving difficulty was greater than 20% being severely penalized by the non-linear index equations. In the linear index formulation some of the bulls with high calving difficulty percentages were being balanced out (at the overall index level) by a high calf value, so the non-linear transformation applies a harsher weighting to these bulls, meaning that they have low index values. While the non-linear index is successful in applying a stronger negative weighting at higher levels of calving difficulty, the trade-off between calf value and calving difficulty at low versus higher levels of calving difficulty is not known. A survey has been designed for Irish farmers to try and quantify the trade-off between calf value and calving difficulty in both dairy cows and heifers, with the strength of the non-linear index weighting that ultimately gets applied in the industry to be determined by the results of the survey.

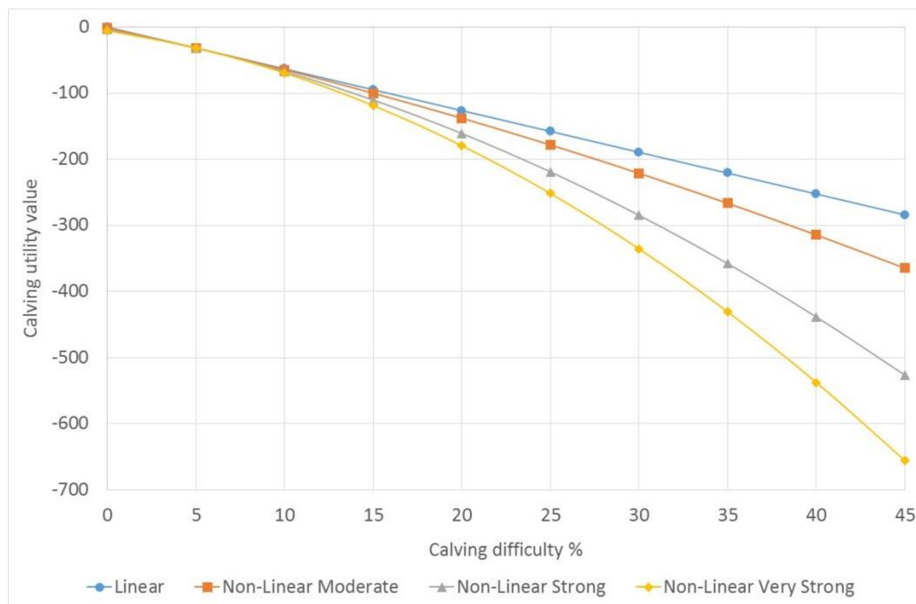


Figure 2. The linear and non-linear index transformations for calving difficulty percent.

REFERENCES

Amer P.R., Simm G., Keane M.G., Diskin M.G. and Wickham B.W. (2001) *Livest. Prod. Sci.* **67**:223

Dekkers J.C.M (1994) *J.Dairy Sci*. **77**: 3441.

Barwick S.A., Henzell A.L. and Graser H-U. (2001) *Proc. Aust. Assoc. Anim. Breed. Genet.* **14**:361.

Berry D.P., Shalloo L., Cromie A.R., Olori V.E. and Amer P. (2005) *ICBF Report*

McHugh N., McParland S., Cromie A.R. and Berry D.P. (2012) *Proc. Agric. Res. Forum.* pp121.

Proc. Assoc. Advmt. Breed. Genet. 21: 13-16

EXTENDED COW LIVEWEIGHT MODELLING FOR BEEF CATTLE BREEDING OBJECTIVES

B.J. Walmsley[1], M.L. Wolcott[1], W.S. Pitchford[2], D.J. Johnston[1] and S.A. Barwick[1]

[1]Animal Genetics & Breeding Unit (AGBU*), Univ. of New England, Armidale, NSW, 2351
[2]Animal and Veterinary Sciences, University of Adelaide, Roseworthy, SA, 5371

SUMMARY

An extended procedure is described for modelling cow liveweight change over the production year for valuing cow liveweight in breeding objectives. The pattern of change, before and after trait change, is able to be approximated from breeder-described variables for any production system. Examination of seven datasets showed cow liveweight changes systematically over age in many herds and breeds. Analyses showed liveweight loss from pre-calving to mating and cow liveweight at mating both change in proportion to the liveweight of the cow pre-calving. Differences in cow liveweight pre-calving are thus larger than are the differences at mating, which affects feed costs in valuing cow liveweight at mating. Adoption of the procedure will increase the precision with which feed cost is associated with cow liveweight in beef cattle breeding objectives.

INTRODUCTION

The breeding female's liveweight is an important breeding objective trait because it affects, especially, the feed needed to maintain the herd or flock. The increase in feed requirement with increasing breeding female liveweight has to be assessed over all parts of the year where feed has a cost. In the BreedObject system for deriving breeding objectives and indexes for beef cattle, cow liveweight in the breeding objective usually has a negative economic value (Barwick and Henzell 2005). The value for cow liveweight is based on modelling liveweight change throughout the year, though that is not very precise. In this study, cow liveweight modelled over the year is extended to better account for the manner in which liveweight loss occurs between pre-calving and mating.

BACKGROUND

Breeding objective traits and values. The breeding objective in BreedObject includes cow weaning rate, cow survival rate, cow liveweight, cow condition score, calving ease (direct and maternal), weaning liveweight (direct and maternal), feedlot entry liveweight, finished sale liveweight, residual feed intake (pasture and feedlot), fat depth, dressing %, carcase meat % and carcase marbling score. Feedlot traits are included when there is a feedlot phase. Cow traits are for an average cow of the herd, with cow liveweight defined at mating (i.e. bull-in). Economic values of a trait are calculated at constant levels of all other breeding objective traits. Included in the economic value of cow liveweight is the cost of the increased feed required for maintenance and liveweight change of the cow. Procedures for assessing feed requirement are given by Freer *et al.* (2007). An increase in cow liveweight also affects returns, as these are greater when surplus cows are heavier.

Breeder description of the production system. Deriving breeding objectives and indexes with BreedObject relies on a breeder description of the production system, where the description is in terms of readily understood variables. Two types of annual feed period affecting cows are recognised: one when feed requirement cannot be increased without there being a cost (i.e. a 'limited pasture' period), and a period (often short) when there would usually be surplus pasture.

---

* AGBU is a joint venture of NSW Department of Primary Industries and University of New England

METHODS

Cow liveweight data. Records were available from an autumn-calving (Struan) Angus herd in southern Australia (Pitchford *et al.* 2015), and from spring-calving Tropical Composite (Belmont-TC, Brian Pastures and Toorak-TC) and Brahman (Swans Lagoon, Belmont-B and Toorak-B) herds in northern Australia (Barwick *et al.* 2009; Wolcott *et al.* 2014). Struan data were available for only 3 parities. Cows in the Struan herd were born in 2 consecutive years and cows in the northern herds were born in 4 consecutive years. Cows in a herd represented 26 to 113 sires. Liveweights in each herd were recorded every 2 to 3 months, including 54 to 60 days pre-calving. Figure 1 shows average weights of cows, in relation to calving and mating, relative to cow age. For convenience, only 4 of the 7 herds are illustrated. Average weight loss between pre-calving (prior to the first calving opportunity) and re-mating (46 to 59 days after calving) in the Struan, Belmont-TC, Brian Pastures, Toorak-TC, Swans Lagoon, Belmont-B and Toorak-B herds was 70.0, 50.8, 66.5, 88.8, 50.1, 54.8 and 67.4 kg, respectively. Corresponding average pre-calving weights were 545.8, 470.6, 454.1, 484.0, 417.2, 450.0 and 464.7 kg.

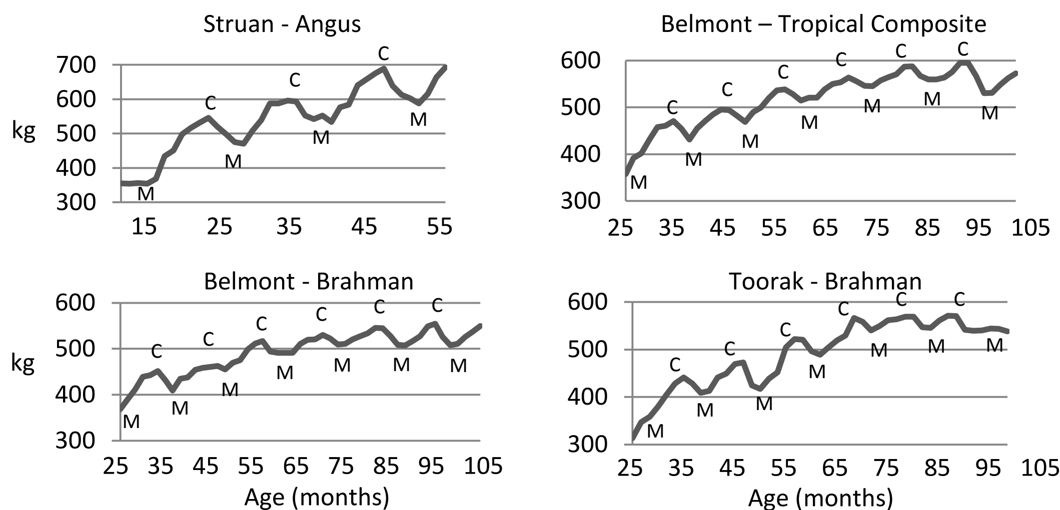


Figure 1. Average cow liveweights (kg), by cow age, for four herds. Also shown are the approximate occurrences of annual calvings (C) and matings (M).

Annual pattern of cow liveweight change. The annual pattern of cow liveweight change can be approximated from the breeder's description of the production system. The illustration in Figure 2 is for an example northern Australian system where the annual dry season is the period of limited feed and the wet season is the period of surplus feed. The system descriptors enabling the pattern to be derived are the timetable of management events, the average liveweight of cows at mating, the average liveweight loss occurring between pre-calving and mating, the length of any time interval pre-mating where the liveweight loss from pre-calving is maximum (shown as 'b' in Figure 2), the length of any period immediately pre-calving where cow liveweight is not increasing (shown as 'a' in Figure 2), and the relative rates of liveweight change between mating and weaning, post-weaning in surplus pasture, and in the early stages of the limited feed period.

Statistical analyses. The liveweight change in cows from pre-calving to re-mating, and mating liveweight, were analysed for each dataset to examine the effect of the regression on pre-calving liveweight. The records included were those from all cows present at the first calving opportunity.

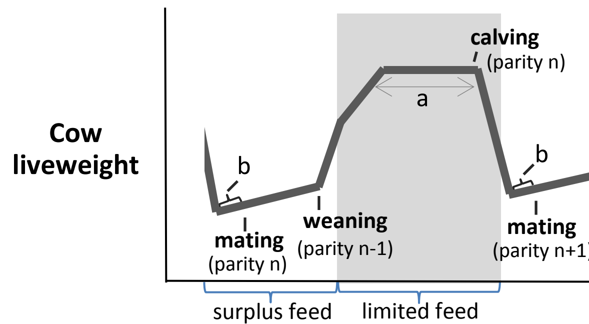


Figure 2. The derived annual pattern of cow liveweight change for an average cow, for an example northern Australian production system (schematic)

SAS Proc Mixed and the R lme procedures were used. The statistical model in each case also included effects that were significant in earlier analyses of the original experiments (Barwick *et al.* 2009; Wolcott *et al.* 2014; Laurence *et al.* 2015). These fixed effects included, for Struan: dam age, year, genotype (Fat vs RFI), genotype x year, High-Fat vs Low-Fat cows within genotype, management group (mg), and age at mating as a covariate nested within genotype (Laurence *et al.* 2015); for Tropical Composite (Brian Pastures, Belmont-TC and Toorak-TC): herd of origin (orig), mg, mg x orig, calf birth month (cbm) x orig (Wolcott *et al.* 2014); and for Brahman (Belmont-B, Toorak-B, Swans Lagoon): mg, cbm, calf sex, cbm x mg (Wolcott *et al.* 2014). For Struan data, replicate within year and herd of origin were fitted as random effects. Sire was included as a random effect in all analyses.

RESULTS

The regression coefficients in Table 1 show cow liveweight loss from pre-calving to mating, and mating liveweight, depend on the cow's liveweight pre-calving. Cows that differed by 100 kg pre-calving were estimated to differ by 8 to 32 kg in their weight loss, on average, across herds and breeds, and by 68 to 92 kg in their mating weight. Differences in cow liveweight, defined at mating, are thus associated with larger liveweight differences at the preceding calving.

Table 1. Regression coefficients (b) for regressions of cow liveweight loss from pre-calving to mating[1], and mating liveweight, on pre-calving cow liveweight in seven herds

Herd and breed	Liveweight loss (kg) from pre-calving to mating			Mating liveweight (kg)		
	b	S.E.	Prob	b	S.E.	Prob
Struan – Angus	0.32	0.06	<.0001	0.68	0.06	<.0001
Belmont - Trop.Comp.	0.20	0.03	<.0001	0.81	0.03	<.0001
Brian Pastures - Trop. Comp.	0.18	0.02	<.0001	0.81	0.03	<.0001
Toorak – Trop. Comp.	0.24	0.04	<.0001	0.77	0.04	<.0001
Swans Lagoon - Brahman	0.08	0.04	0.0724	0.92	0.04	<.0001
Belmont - Brahman	0.19	0.03	<.0001	0.81	0.03	<.0001
Toorak - Brahman	0.13	0.06	0.0316	0.87	0.06	<.0001

[1]Liveweight loss between a cow's first calving opportunity and the subsequent mating

The cow liveweight curve that applies after a change in cow liveweight at mating (ie. at parity n+1) (Figure 3) is able to be approximated using the same variables. The relationship that is used

between cow liveweight loss from pre-calving to mating and pre-calving liveweight is taken from the breeder description and thus is specific to the production system described.

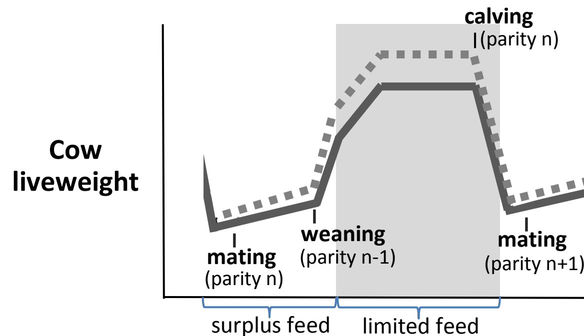


Figure 3. The derived annual pattern of cow liveweight change, before (▬) and after (▪▪▪) change in cow weight, for an example northern Australian production system (schematic)

DISCUSSION AND CONCLUSIONS

Figure 1 illustrates the systematic annual pattern of cow liveweight change that occurs in many herds and breeds. It supports the description given by Cole (1970). The annual pattern for an average cow, before and after change in cow liveweight, can be approximated from breeder-described variables for any production system (Figures 2 and 3).

Table 1 shows cow liveweight loss from pre-calving to mating, and cow weight at mating, change in proportion to the liveweight of the cow pre-calving. A change in cow liveweight at mating is consequently associated with a greater change in cow liveweight pre-calving (Figure 3). This affects the value of cow liveweight in the breeding objective, as the greater requirement pre-calving usually affects feed costs. It is the difference in feed requirement over all periods of the year when feed has a cost, before and after trait change, that determines the feed cost associated with the change.

The described procedure has been incorporated in BreedObject and will improve the precision with which cow liveweight is linked to feed cost in deriving individual breeding objectives.

ACKNOWLEDGEMENTS

We acknowledge the contributions to datasets made by participants in the beef industry CRCs, and thank NSW Dept. of Primary Industries and Meat & Livestock Australia for financial support.

REFERENCES

Barwick S.A. and Henzell A.L. (2005) *Aust. J. Exp. Agric.* **45**: 923.

Barwick S.A., Johnston D.J., Burrow H.M., Holroyd R.G., Fordyce G. et al. (2009) *Anim. Prod. Sci.* **49**: 367.

Cole V.G. (1970) 'Beef production guide'. Winn & Co. Marrickville, Graziers' Assoc. of NSW.

Freer M., Dove H. and Nolan J.V. (Eds)(2007) 'Nutrient requirements of domesticated ruminants'. CSIRO Publishing, Collingwood Vic.

Laurence M., Accioly J.M., Copping K.J., Deland M.P.B., Graham J.F. et al. (2015) *Anim. Prod. Sci.* **55** (in press)

Pitchford W.S., Accioly J.M., Banks R.G., Barnes A.L., Barwick S.A. et al. (2015) *Anim. Prod. Sci.* **55** (in press)

Wolcott M.L., Johnston D.J., Barwick S.A., Corbet N.J. and Williams P.J. (2014) *Anim. Prod. Sci.* **54**: 37.

Proc. Assoc. Advmt. Breed. Genet. 21: 17-20

EVALUATION OF ALTERNATIVE SELECTION INDEXES FOR NON-LINEAR PROFIT TRAITS APPROACHING THEIR ECONOMIC OPTIMUM

D. Martin-Collado[1], B. Visser[1], T. Byrne[1], and P. Amer[1]

[1]AbacusBio Limited, Dunedin 9016, New Zealand

SUMMARY

This paper presents a simulation which evaluates the performance of alternative selection index configurations in the context of a breeding program where traits with non-linear economic value are approaching an economic optimum. The simulation described uses a simple population structure that approximately mimics selection in dual purpose flocks in New Zealand, where number of lambs born is believed to be approaching an optimum, while other genetically correlated traits with linear economic values are assumed to not be approaching any economic optimum. A "non-linear below the optimum and then flat" approach to index formulation was found to be at least comparable in efficiency to the approach of regularly updating the linear index with short (15 year) and long (30 year) time frames, especially when the current average value of the "non-linear" trait is at a reasonable distance from the optimum. Use of a non-linear index that is efficient may have other benefits in highly heterogeneous industries (breeds and production environments) such as the New Zealand sheep industry.

INTRODUCTION

Number of lambs born (NLB) is a key trait that has delivered significant economic value from genetic improvement to the New Zealand sheep industry to date. However, its genetic improvement might be reaching an economic optimum above which further increases are not desirable, because of an increase in costs that would decrease the profitability of animals. There is therefore a need to define a selection index which could hold traits such as NLB at their optimum levels, while focusing selection on other traits which are not approaching any economic optimum.

The predominant view among theoretical livestock geneticists, that has not been revised for a number of years, is that when faced with a non-linear profit function for one or more traits, the best approach is to still use linear selection indexes as they are optimal when regularly updated (Goddard 1983; Meuwissen and Goddard 1997; Dekkers and Gibson 1998). However, when the trait average is near the optimum or when profit functions are extremely non-linear, linear selection indexes would be unsatisfactory (Goddard 1983; Dekkers and Gibson 1998). Furthermore, Meuwissen and Goddard (1997) showed that non-linear indexes can approach the response achieved by linear indexes while not requiring any updating. Nevertheless, there are more options which, to our knowledge, have never been tested: for example use of a non-linear index before the optimum and then "flat", i.e. the marginal economic value is assigned a value of zero.

The aim of this study is to the compare the effect of alternative selection index approaches on the genetic change in a trait with non-linear economic value, approaching its economic optimum, and traits with linear economic value. The efficiency of the selection indexes is measured in terms of genetic progress achieved in the population and in economic benefits achieved in the short, medium, and long-term.

METHODS

The model simulates 30 generations of a population of 1000 males and 1000 females. Two genetically correlated (0.07) traits were simulated: one trait with a non-linear economic value (T_{NL}, h^2=0.1) and the other one with a linear economic value (T_L, h^2=0.25). Genotypes for the animals in

the base population were simulated from a random normal distribution $N\ (0, \sqrt{h^2})$. Phenotypes and genotypes of the base population and subsequent generations were simulated using standard approaches as described by Hely *et al.* (2012). In each generation estimated breeding values (eBVs) were calculated by applying alternative selection index functions (described below) to their phenotype. The best 20% of the males born in each year are selected to become sires in the next generation. Conversely, for dams an aging process was simulated. In the base population females were assigned randomly to one of three cohorts (age groups). In each year the oldest dam cohort was culled and replaced by the best females born from the previous year. Dams and sires were randomly mated. One offspring was simulated for each unique parent mating type. Then, the economic performance of the selection approach was calculated by applying the profit functions (described below), to the average population phenotypic value of T_{NL} and T_L. The economic performance is given as discounted profit to express the profit of future generations at present value with a discount rate of 0.07 per year.

Profit functions. Profit functions defined the true economic merit of the individuals and were used to quantify the economic performance of alternative selection approaches at population level. They consisted of a linear function for T_L and a non-linear function (quadratic) for T_{NL}. Two profit functions were defined which differed in the distance of the T_{NL} optimum to the initial population T_{NL} average (T_{NL}=0): (1) a "close to the optimum" profit function that had the optimum at T_{NL}=2, and (2) a "distant to the optimum" function that had the optimum at T_{NL}=4. T_{NL} optimum values where arbitrarily defined so that the optimum values were reached in the time frame considered (30 years) when applying some of the selection index functions evaluated.

Selection index functions. The selection index functions always gave T_L a constant linear weighting. The weighting approach for T_{NL} defined the four alternative selection index functions evaluated:

Linear index. The T_{NL} component is linear, with the linear slope value calculated as the partial derivative of the profit function at the initial population mean.

Linear index updated periodically (LUP index). The T_{NL} component is linear but the slope of the linear function is updated (each 3 or 5 years) to match the slope of the T_{NL} non-linear profit function being used.

Non-linear index. The T_{NL} component is non-linear and is identical to the corresponding T_{NL} profit function being used.

Non-linear then flat index (NLTF index). Before the optimum, the T_{NL} component is non-linear and is identical to the corresponding T_{NL} profit function being used. After the optimum the marginal economic value takes a value of zero. Thus, animals with eBVs below the optimum are penalised, while animals with eBVs at or above the optimum are not penalised.

RESULTS AND DISCUSSION

Table 1 presents the economic performance of the alternative selection indexes evaluated with the two profit functions for selecting for T_L and T_{NL}. Figure 1 shows the evolution of the population average T_L and T_{NL} phenotypic values and of the discounted profit when the profit function is set so that the T_{NL} optimum was distant from the initial population value.

Linear selection indexes are initially the fastest way to increase average T_{NL} and reach the optimum but since they continue selecting above the optimum they ultimately become counterproductive. This selection pressure on T_{NL} reduces the selection space for T_L which is therefore not heavily selected for.

LUP indexes are more efficient than Linear indexes since while they still achieve a fast improvement of T_{NL} before the optimum, once the population average reaches the optimum they stop selecting for T_{NL}. However, since the same index is applied to all the individuals in the

population, once the population average reaches the optimum, individuals below the optimum are not penalised which results in a loss of some selection potential for T_{NL}.

Non-linear selection indexes decrease the rate of response as the population gets close to the optimum and negatively select those individual above the T_{NL} optimum. Therefore, the non-linear selection indexes do not allow the population to surpass the optimum. Animals at either extreme from the optimum are not considered as selection candidates, even though with relatively balanced selection of animals at both extremes, the population mean would not move away from the optimum. These effectively excluded candidates are not considered for selection on T_L and therefore the potential selection on T_L is not fully realized. The selection pressure on the non-linear trait, when close to the optimum, is not very intense so the rate at which the non-linear index approaches the optimum is slower than linear selection indexes.

A **NLTF index** achieves a selection speed before the optimum intermediate to the linear selection indexes and the pure non-linear indexes, leaving more selection space to the T_L than linear selection indexes. After the optimum the NLTF keeps on penalizing for T_{NL} all those animals below the optimum, while those animals above the optimum are not negatively selected for T_{NL}, allowing them to be selected for T_L.

Table 1. Cumulative discounted profit achieved by applying alternative selection indexes

Trait with non-linear profit function	Period (years)	Selection indexes				
		Linear	**Linear 3-year update**	**Linear 5-year update**	**Non-linear**	**Non-linear then flat**
Close to optimum	1-10	12.4	14.9	14.2	10.1	**15.8**
	1-15	16.2	26.6	24.9	18.4	**27.4**
	1-20	13.4	**38.1**	35.1	26.4	26.4
	1-25	4.6	**48.4**	44.1	33.5	44.7
	1-30	-8.5	**57.2**	51.7	39.4	50.5
Distant from optimum	1-10	16.6	17.0	17.0	14.8	**18.1**
	1-15	28.1	30.6	30.4	26.8	**32.7**
	1-20	36.4	43.5	43.1	38.2	**46.2**
	1-25	41.1	55.1	54.5	48.2	**57.8**
	1-30	42.3	64.9	64.3	56.5	**67.3**

CONCLUSION

NLTF selection indexes are the most optimal indexes to select for traits with non-linear economic values in the short and mid-term when the average population is relatively close to the optimum (Table 1). After that period it becomes less profitable than the updated linear selection indexes because the negative profit, due to the T_{NL} average population value being far above the optimum, is not offset by the better selection for T_L. However, when the optimum profit is set to be relatively distant from the current T_{NL} average population value, the NLTF selection index is the most profitable of all the indexes assessed for the time span evaluated. There could well be other advantages of the NLTF selection approach. For example, a single index of this makeup could be applied across a wide diaspora of breeds and flocks differing in their current level of merit for the non-linear trait. In the New Zealand sheep context, breeder flocks with low average genetic merit for NLB would have a high weighting applied to NLB, while those with average merit at or beyond the economic optimum would not. However, there are also some challenges for the implementation of an index with a NLTF selection function. The simple multiplication of eBVs by known economic weights, to produce a selection index, is lost for the non-linear trait

when NLTF selection approach is applied. This has the potential to create a communication and extension challenge for index users. Consideration will also need to be given to the base value for the non-linear index.

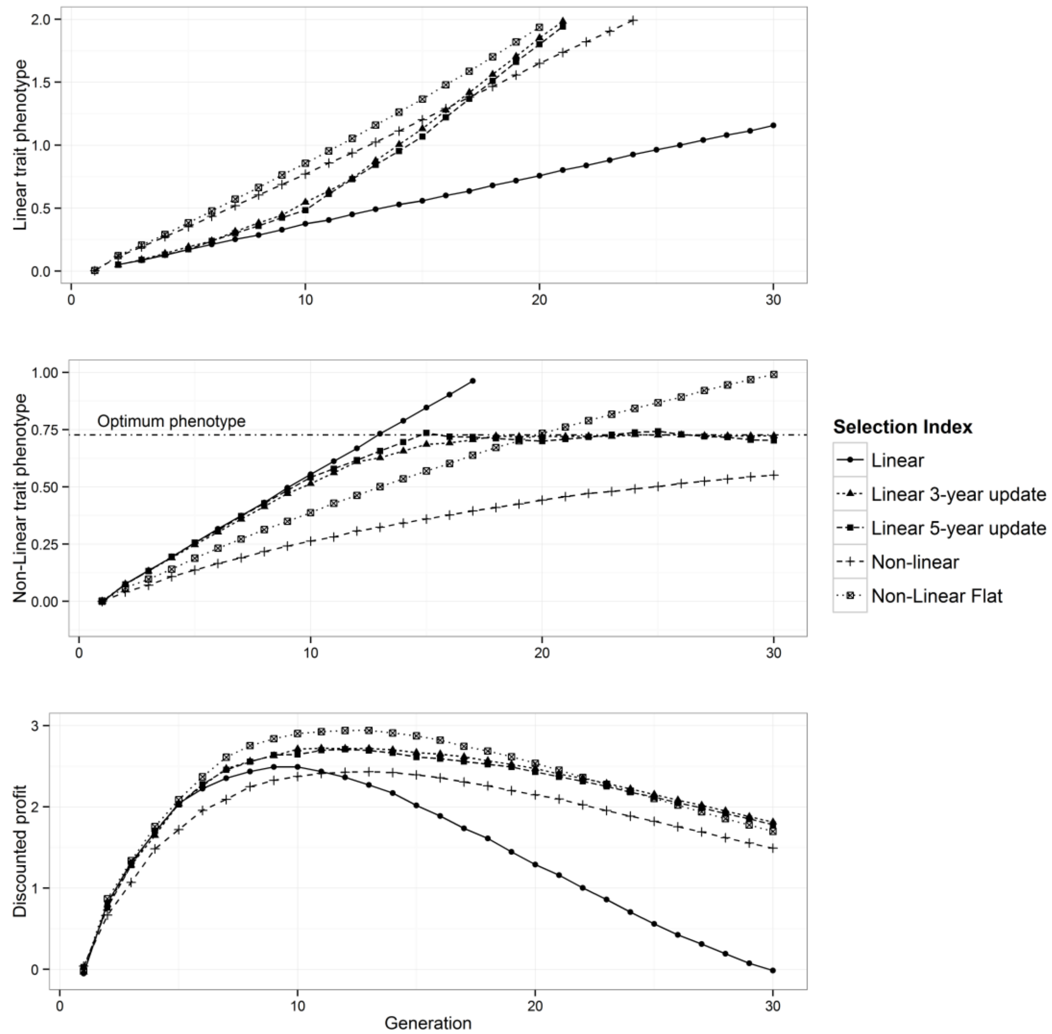


Figure 1. Evolution of average population phenotypic value of the traits with non-linear and linear economic value when selected on the alternative selection indexes. Case of profit function with distant optimum for the non-linear trait.

REFERENCES

Dekkers J. and Gibson J. (1998) *J. Dairy. Sci.* **81**: 19.
Goddard M. (1983) *Theor. Appl. Genet*. **64**: 339.
Hely F.S., Amer P.R., Walker S.P. and Symonds J. E. (2012). *Animal* **7**: 1.
Meuwissen T.H.E. and Goddard M. (1997). *Anim Sci.* **65**: 1.

Proc. Assoc. Advmt. Breed. Genet. 21: 21-24

A REVIEW OF THE NATIONAL BREEDING OBJECTIVE AND SELECTION INDEXES FOR THE AUSTRALIAN DAIRY INDUSTRY

T. Byrne[1], D. Martin-Collado[1], B. Santos[1,2], P. Amer[1], J. Pryce[3], and M. Axford[4]

[1]AbacusBio Limited, Dunedin 9016, New Zealand
[2]School of Environmental & Rural Science, University of New England, Armidale, Australia
[3]Department of Economic Development, Jobs, Transport and Resources and La Trobe University, Bundoora, VIC 3083, Australia
[4]Australian Dairy Herd Improvement Scheme, Melbourne VIC 3000, Australia

SUMMARY

We applied a pairwise comparison method using the 1000Minds® software to assess farmers' preferences for cow trait improvements. A Principal Component Analysis (PCA) followed by a Cluster Analysis (CA) of the principal components led to the identification of three farmer clusters (farmer types in the rest of this document) according to the trait improvements to which the farmers had the highest preference. This way, Australian dairy farmers can be classified into production-focused (n = 192), functionality-focused (n = 187), and type-focused (n = 172) farmers. As a result of this study, and bio-economic modelling, three indexes were released to the Australian dairy industry. The Balanced Performance Index aligns with the average preferences, while the Health Weighted and Type Weighted indexes reflect the preferences identified for functionally-focused and type-focused farmer types, respectively. These three indexes include new traits and offer a range of options to choose from when selecting bulls, while all driving gain towards the National Breeding Objective (NBO).

INTRODUCTION

Breeding objectives can play an important, but not exclusive, role in determining the optimal size and direction of genetic changes in traits. Economically efficient multiple-trait selection is normally achieved through the definition of breeding objectives and the development of appropriate selection indexes for specific production systems (James 1981). In nations with industrialised dairy industries a breeding objective is often controlled at the national level (e.g. Harris *et al.* 1996). The NBO underpins the selection index for the ranking of dairy cattle for profitable genetic merit in Australia (Pryce *et al.* 2010). The aim of this study was to update the NBO by calculating economic weights for a range of traits that impact profitability of Australian dairy farms. The final choice of selection indexes was informed by analysing the heterogeneity of farmers' preferences (from surveys) for improvements in dairy cow traits using farmer typologies.

This paper broadly describes the methodology used to analyse heterogeneity of farmers' preferences and how the outcomes of this were used, along with economic analysis underpinning the breeding objective, to develop selection indexes.

METHODS

Survey questionnaire and analysis. We applied a pairwise comparison method to assess farmers' preferences for trait improvements, using the 1000Minds® software. This software is simple to implement and reduces the level of burden on respondents compared to other more complex methods (Hansen and Ombler 2009). The software asks a series of questions to respondents, who are asked to choose, repeatedly, between pairs of alternatives until all possible pairs of alternatives are evaluated. A ranking of the presented alternatives is derived from these choices. We considered most of the traits included in the Australian Profit Ranking (APR), at the time of

surveying, as well as other traits that were considered of potential importance for the Australian dairy industry. Survey traits included; protein yield, cow live weight, fertility, longevity, mastitis resistance, milking speed, temperament, calving difficulty, feed efficiency, lactation persistency, lameness, mammary system, and overall type. The magnitude of the suggested improvement in each trait was such that our estimate of the economic impact on farm would be as similar as possible across traits (Martin-Collado et al. 2015). Farmer attitudes towards genetic evaluation tools were assessed by asking farmers to rate, in a five-level Likert scale (Likert 1932), their level of agreement with specific statements. Farmers were also asked a set of farmer and farm descriptors that were thought to have a potential influence on farmers' preferences for improvements in traits. These included farmer age, role on farm, farm location, herd size, total milk production, cow breed distribution, cows registered with breed society, replacements sired by AI or herd bulls, labour profile, calving system, and feeding system. Farmers of all 6314 Australian dairy farms were sent the survey. In addition, 200 levy-paying farmers were randomly selected from the list of all Dairy Australia farmers. The survey produced 618 responses, of which 551 were fully completed and were used for this study.

A Principal Component Analysis (PCA) followed by a Cluster Analysis (CA) of the principal components was used to investigate the patterns of relationships between farmers' preferences for the different trait improvements. We determined the principal components (PCs) of the trait preferences and implemented a Ward's Hierarchical CA of the first five principal components. The selection of the number of clusters was based on the loss of inertia (within cluster sum of squares) at each partitioning of clusters (Ward 1963). We described the farmer types according to their preferences for animal trait improvements. We analysed the relationship between farmer types and farmer attitudes, criteria used for selecting bulls (results not shown) and other farm and farmer descriptors (as reported above). Differences for the normally distributed variables were analysed with the ANOVA test followed by Duncan's multiple comparisons test to analyse pairwise differences. The non-normally distributed variables were analysed with the Kruskal-Wallis test and multiple comparisons were tested with the Wilcoxon's procedure. Finally, the Fisher's exact test was used to analyse pairwise differences between discrete variables among farmer types.

Formulation of breeding objectives and selection indexes. Economic weights in the breeding objective were calculated as the economic effect on profit per unit change in each of the traits independently, allowing for the Australian dairy production system diversity of feeding systems and calving patterns. These economic weights are reported elsewhere (Byrne *et al.* in preparation). Selection indexes were defined using a combination of economic principles and desired gains approaches, such that indexes remained relevant for improving on-farm profit based on strong scientific principles which were also consistent with farmers' preferences.

RESULTS AND DISCUSSION

In the overall ranking of preferences for trait improvements at population level we could distinguish the most preferred and the least preferred trait improvements, as well as a large number of trait improvements with medium preference. Mastitis (average rank 4.3) was the most preferred trait followed by longevity (5.1) and fertility (5.4) whereas the least preferred traits were milking speed (8.2), lactation persistency (8.3), and cow live weight (10.4). These preferences are relative to crude calculations that equalise the economic effects of each offered trait difference; thus the preferences are more likely to be driven by perception than by economics.

Principal Component Analysis of Farmers' Preferences for Trait Improvements. The scores of farmers' preferences for trait improvements in the first two PCs are described in Figure 1. These

first two PC accounted for 26.6% of the total variability of the farmers' trait improvement preferences, and five PCs were needed to explain 55.5% of the initial variability.

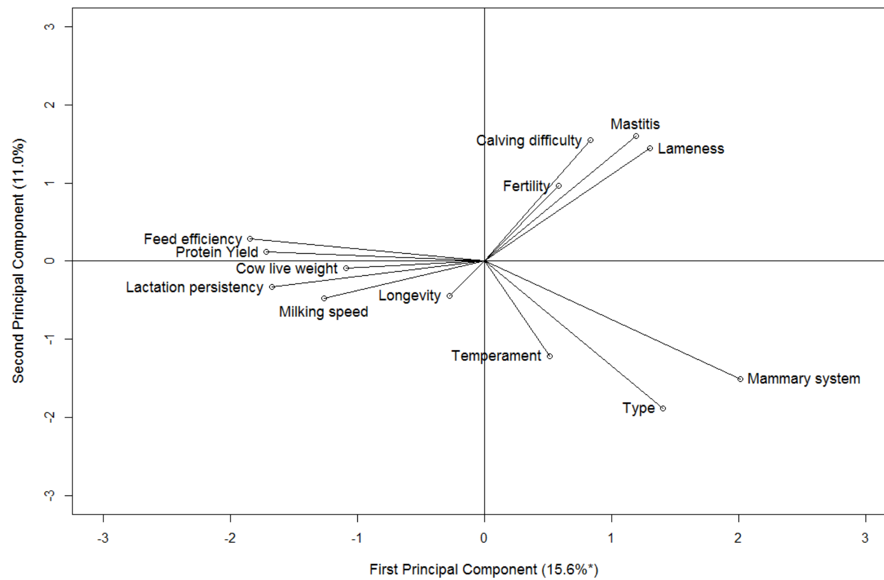


Figure 1. Scores of the preferences for improvements on cow traits on the first two principal components.

Cluster Analysis of the principal component. While the data indicates a continuum of preference, the cluster analysis of the first five PCs determined the existence of three farmer types of very similar sizes, named according to the trait improvements to which the farmers had the highest preference. This way, Australian dairy farmers can be classified into production-focused (n = 192), functionality-focused (n = 187), and type-focused (n = 172) farmers.

Production-focused farmers gave the highest preference to improving longevity (mean rank±SE: 4.4±0.23), feed efficiency (5.2±0.22), and protein yield (5.3±0.23). Compared to the other farmer types production-focused farmers gave the highest importance of all to protein yield, lactation persistency (6.3±0.25), feed efficiency, cow live weight (9.0±0.25), and milking speed (6.9±0.26). Conversely, they gave lowest importance of all the farmer types to improving mastitis (5.8±0.27), lameness (8.1±0.23), and mammary system (8.4±0.21).

Functionality-focused farmers gave the highest preference to mastitis (2.8±0.17), followed by lameness (4.6±0.26), calving difficulty (5.2±0.22), and fertility (5.4±0.25). Compared to the other farmer types, functionality-focused farmers gave the highest preference of all to mastitis, lameness, and calving difficulty.

Type-focused farmers preferred improvements in mammary system (3.7±0.15), longevity (4.0±0.19) and mastitis (4.1±0.20) the most. Compared to the other farmer types, type-focused farmers gave the highest preference of all to mammary system, and type (4.9±0.19). On the contrary, type-focused farmers gave the lowest importance of all to protein yield (8.5±0.22).

There was an expectation that factors such as farm size and calving or feeding system would explain some of the variability in farmers' preferences for trait improvements, but we did not find significant differences between farmer types for any of the farm descriptors. However, in a univariate analysis of the survey results, we observed that the importance given to specific traits

was related to some of the farm features. Seasonal calving farmers gave higher preference (ANOVA p-value < 0.05), average rank 4.9, to an improvement in cow fertility compared to farmers of split-calving herds (5.5) and all-year-round herds (5.8) and to not increasing live weight (ANOVA p-value < 0.001), average rank 9.6, compared to the other calving systems (pooled average of 10.7). There was also no clear relationship between farmers' preferences and breed when analysing the PC clusters. The results could imply that farmers' preferences are intrinsic to the farmer, rather than being strongly linked to external system factors.

Formulation of selection indexes informed by farmers' preferences. Australian dairy farmers can be divided into three types according to the pattern of their preferences for trait improvements. As a result of detailed bio-economic modelling, and this study, three indexes were released to Australian dairy farmers (Figure 2) in September 2014. These three indexes include new traits, informed by trait preference data, and offer a range of options to choose from when selecting bulls. The Balanced Performance Index aligns with the average preferences, while the Health Weighted and Type Weighted Indexes reflect the preferences identified for Functionally-focused and Type-focused farmer types, respectively. The economic weights for all traits were calculated based on economic principles, with the exception of a number of trait weightings in the Type-weighted index, which were calculated using a desired gains approach informed by trait preference data.

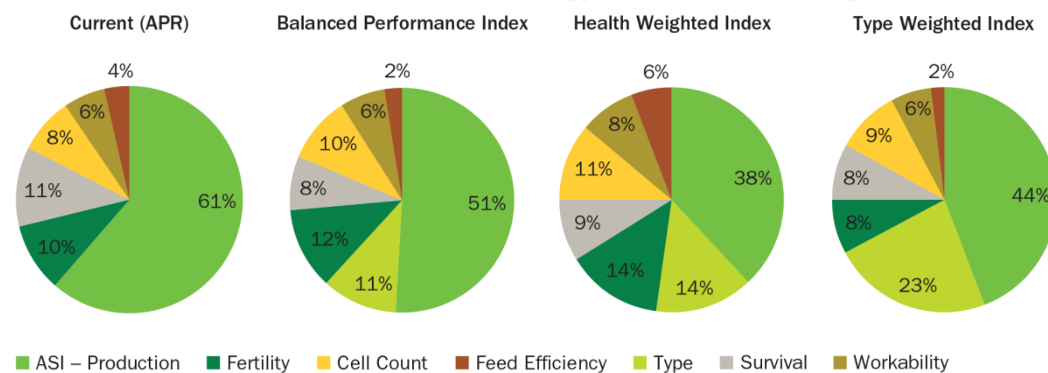


Figure 2. Relative emphasis in the three new indexes and the APR.

CONCLUSION

There are different groups of Australian dairy farmers with specific needs. This has led to the three indexes including new traits and offers a range of options when selecting bulls.

REFERENCES

Hansen P. and Ombler F. (2009) *J. Mult. Crit. Dec. Anal.* **15:** 87.

Harris B., Pryce J.E. and Montgomerie W. (1996). *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17**: 434.

James, J.W. (1981). In D. K. Belyaer, (ed.), *Proc. 14th Int. Con. Genet*. Moscow, pp. 221.

Likert R. (1932) *Arch. Psychol.* **22**: 5.

Martin-Collado D.**,** Byrne T.J., Amer P.R., Santos B.F.S., Axford M. and Pryce J.E. (2015). *J. Dairy Sci.* **98**: 4148.

Pryce J.E., van der Werf J.H.J., Haile-Mariam M., Malcolm B. and Goddard M. (2010) A technical manual describing the Australian Profit Ranking (APR Index) - Version 2.

Ward J. H. (1963) *J. Am. Stat. Assoc.* **58**: 236.

Proc. Assoc. Advmt. Breed. Genet. 21: 25-28

FEED SAVED BREEDING VALUES FOR AUSTRALIAN DAIRY CATTLE

J.E. Pryce[1,2], O. Gonzalez-Recio[1], G. Nieuwhof[1,3], W.J. Wales[4], M.P. Coffey[5], B.J. Hayes[1,2], and M.E. Goddard[1,6]

[1] AgriBio, Economic Development and Dairy Futures CRC, VIC 3083, Australia
[2] School of Applied Systems Biology, La Trobe University, VIC 3083, Australia
[3]Australian Dairy Herd Improvement Scheme, 22 William Street, Melbourne, VIC 3000, Australia
[4] Economic Development, Ellinbank, VIC 3820, Australia
[5]SRUC, Easter Bush Campus, Midlothian EH25 9RG, UK
[6]Faculty of Vet. and Agric. Sciences, The University of Melbourne, Parkville,VIC 3010, Australia

SUMMARY

A new breeding value is described that includes the amount of feed saved per year through assumed improvements in lifetime metabolic efficiency and reduced maintenance requirements. The breeding value includes a genomic breeding value for residual feed intake, which is available for Holsteins only, combined with either genomic or pedigree estimated breeding values for maintenance requirements predicted using type traits. The standard deviation of estimated breeding values for feed saved was 65.6 kg/year. The mean reliability of feed saved was 0.37 and the standard deviation was 0.06, with the range from 0.24 to 0.74. The breeding value, which will be known as "feed saved", has been available for dairy farmers and breeding companies to use in their selection decisions since April 2015.

INTRODUCTION

While it is widely recognised that selecting for feed efficiency in dairy cattle is highly desirable, as feed costs comprise a large proportion of variable costs, there has been little success in developing breeding values to select directly for this trait in dairy cattle breeding (Berry and Crowley, 2013).

Genomic selection is well suited to traits that are measured in small, well recorded populations. Residual feed intake (RFI) is one such trait that fits this description when measured on genotyped individuals. The reference population can be used to develop genomic prediction equations that can then be applied to genotyped animals without phenotypes.

Here, a breeding value called feed saved is proposed. Feed saved is made up of genomic breeding values for RFI of cows and calves combined with maintenance requirements predicted from type breeding values, so that feed requirements are quantified in a single breeding value. For example, where 2 individuals have similar breeding values for milk production traits, it becomes obvious that the animal requiring less feed (for maintenance predicted from BW and RFI combined) will be more efficient and save more feed, all other things being equal.

The aim of this paper is to describe how breeding values could be calculated and implemented for this new trait.

MATERIALS AND METHODS

The process to estimate genomic breeding values (GEBVs) for feed saved includes: 1) calculating phenotypes for RFI of animals included in the reference population; 2) calculating direct genomic values (DGVs) for RFI in the growth and lactation stages of life, and 3) calculation of feed saved GEBVs by combining body weight (BW) EBVs with RFI DGVs.

The reference population used for genomic prediction of RFI was similar to the population used by de Haas *et al.* (2012) and Pryce *et al.* (2014), except additional Australian lactating cows

were included and RFI was the phenotype instead of dry matter intake (DMI).

The residual feed intake of 843 growing calves (RFIcalf) was calculated from means of DMI measured over a 6 to 7 week period in growing Holstein calves of around 6 months of age, regressed on average BW and growth over the experimental period (Williams *et al.* 2011).

Phenotypes for RFI in Australian lactating cows (RFIcow) were calculated from means of DMI measured over a 30 day period in multiparous lactating cows regressed on average BW and energy corrected milk.

The UK and Dutch data reported by de Haas *et al.* (2012) were used to calculate RFI in overseas cows (RFIov), and included DMI phenotypes pre-corrected for fixed effects and regressed on energy corrected milk calculated from GEBVs for milk production traits and BW. These GEBVs were calculated using Australian Dairy Herd Improvement Scheme's (ADHIS) official genomic prediction equations, as phenotypes for these overseas animals were not available.

There were 28,621 SNPs in common between Australian and overseas datasets and the SNPs were very similar to those used in the study of de Haas *et al.* (2012). Briefly, the Australian calves were genotyped using the Illumina High Density Bovine SNP chip (Illumina, San Diego, CA; www.illumina.com/agriculture), after quality control procedures described by Pryce *et al.* (2014), 624,930 SNP remained. The Australian, UK and Dutch cows were genotyped using the Bovine SNP50 Beadchip (Illumina Inc., San Diego, CA) and were edited as described by de Haas *et al.* (2012).

A multi-trait analysis (RFIcalf, RFIcow, RFIov) was used to generate GEBVs for RFI in Australian calves and cows, and in overseas cows. This model allowed for the correlations between traits to be estimated.

The model used was:

$$y_T = X_T b_T + Z_T g_T + e_T$$

Where $\mathbf{y_T}$ is the 3xn matrix of observations on all traits, $\mathbf{X_T}$ is the incidence matrix for fixed effects, $\mathbf{b_T}$ is the matrix of solution of fixed effects (in this case the mean for each trait and country of origin), $\mathbf{Z_T}$ is an incidence matrix mapping records to animals, $\mathbf{g_T}$ is the corresponding genomic breeding values for animals with genotypes for all traits, and $\mathbf{e_T}$ is a 3xn matrix of residual terms. Variance components were estimated with ASREML (Gilmour, 2006).

Regression equations that allow both RFIcalf and RFIcow to be computed from SNP markers were calculated by back solving the mixed model equations. The estimation of SNP coefficients was calculated as $\beta = H'(HH')^{-1}\hat{g}$, where H is the n x 28,621 matrix of the genotypes of n animals in the reference set, with genotypes coded as 0, 1, 2 for aa, Aa and AA respectively, and $\hat{g}$ is the DGV for RFIcalf or RFIcow.

RFIcalf and RFIcow DGVs were combined to calculate lifetime RFI (RFIlife). RFIlife was expressed in kg of feed per year and it was assumed that the rearing period is 2 years and cows have 4 lactations on average. This was done by multiplying the DGV for RFIcow by days in lactation (days), which was assumed to be 305d and deliberately excluded the dry period when feed costs and daily feed intake were lower (than during lactation). Similarly, RFIcalf was multiplied by days of life before lactation commenced, which was assumed to be 700d (i.e. ignoring the period from birth to weaning, which is approximately 30d, when the diet is predominantly milk). Finally, there is a division by 6 (2 years as a calf + 4 years as a lactating cow).

The feed saved breeding value is defined as the annual feed required for maintenance combined with residual feed intake. Maintenance requirements (in MJME) are generally considered to be a function of BW (e.g. Visscher *et al.* 1994). In Australia, BW breeding values are calculated from type traits (see Haile-Mariam *et al.* (2014) for more details). The marginal change in feed required to maintain a given EBV of BW was calculated. Finally, the direction of

the EBV was reversed, so that high values meant that more feed was saved.

The reliabilities of BW EBVs are already calculated routinely as part of the genetic evaluation service of the ADHIS. A method to estimate the reliabilities for RFIcalf and RFIcalf for genotyped animals without phenotypes was also required. This was achieved by inverting the coefficient matrix that incorporates animals with and without records in the system of equations.

RESULTS AND DISCUSSION

The rationale for combining RFI and feed required for maintenance predicted from BW is that annual feed requirements are accounted for in the same breeding value. Selecting for this trait in tandem with other traits of economic importance via a selection index, leads to selection for improved feed efficiency. The advantage of feed saved is that animals with the same RFI can be distinguished on the basis of their maintenance requirements. Then, everything else being equal, larger animals will be penalised for greater maintenance costs. An important distinction and the reason why feed saved (integrating RFI and maintenance costs associated with higher BW) is a more desirable trait for implementation than RFI, is because RFI is corrected for BW.

Table 1. Phenotypic SD, heritability (diagonal) and genetic correlation estimates for RFI measured in Australian cows (RFIcow), Australian calves (RFIcalf) and UK and Dutch cows (RFIov) (above the diagonal). Standard errors are presented in parentheses.

	Phenotypic SD	RFIcow	RFIcalf	RFIov
RFIcow	1.26	0.20 (0.20)	0.67 (0.45)	0.76 (0.60)
RFIcalf	0.42		0.35 (0.08)	0.30 (0.22)
RFIov	0.97			0.35 (0.06)

The heritability of RFI in Australian cows, estimated using genomic relationships between animals, was 0.20 and for growing calves and a combined dataset of Dutch and UK cows, respectively, was 0.35 (Table 1), both of which are in the range of previous studies (e.g. Berry and Crowley, 2013). The standard errors of the estimates were large in general, as expected given the limited data, especially for Australian lactating cows. As more Australian lactating cow data accumulates, the standard errors are expected to reduce, which would lead to less reliance on data from other countries.

In addition to a small number of Australian cow records for RFI, the other limitation in how RFI has been evaluated in this study, was a failure to account for differences in body composition or changes in body composition. Although body condition score (BCS) is the accepted measure of fat reserves in dairy cattle, it is generally scored on a relatively limited scale which makes evaluating changes in BCS difficult for a dataset of only several hundred cows over a relatively short time period. Advancements in accurately quantifying body reserves in dairy cattle will benefit genetic prediction of RFI. For the time being, including feed saved as part of a multi-trait selection index for profit, that also includes fertility as well as traits like survival and mastitis resistance, should help to guard against undesirable correlated responses, for example in fertility.

The DGVs for RFIcalf and RFIcow had standard deviations of 0.09 kg/d and 0.21 kg/d respectively. The standard deviations of estimated breeding values for RFIlife and feed saved were 50.5 and 65.6 kg/year. The mean reliability of feed saved was 0.37 and standard deviation was 0.06, with the range from 0.24 to 0.74.

The genetic trend for feed saved is shown in Figure 1 and includes a subset of the genotyped bulls that were born from 1990 onwards. It can be seen that from around 2000 there has been an

increase in breeding values for maintenance requirements and RFI, consequently the genetic trend for feed saved is unfavourable.

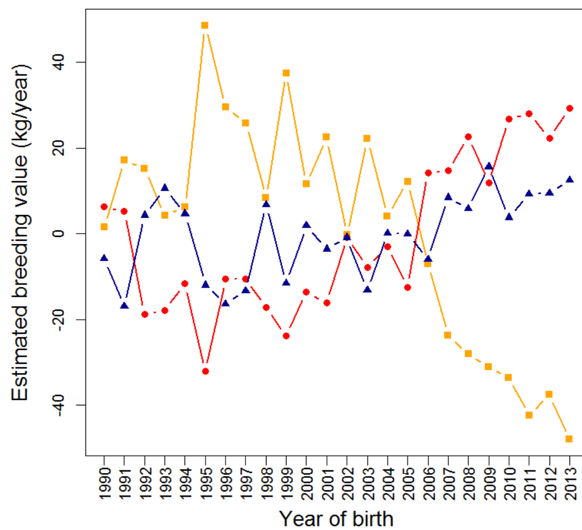


Figure 1. Genetic trend of estimated breeding values for bulls born in or after 1990 for RFIlife (▲), annual feed required for maintenance (●) and Feed Saved (■).

To conclude, selection for feed efficiency in dairy cattle historically has relied on approximations of maintenance requirements. However, this is unlikely to capture all the variation in feed efficiency. Through genomic selection, there are now opportunities to extend this to include measures of efficiency derived from actual feed intake data, such as RFI. This study described how a feed efficiency breeding value can be derived by combining RFI with maintenance requirements. Although the mean reliability of this new trait was comparatively low when compared to other traits, this is expected to improve as the reference population is expanded.

REFERENCES

Berry, D. and Crowley J. (2013) *J. Anim. Sci*. **91**:1594-1613.

De Haas, Y., Calus M, Veerkamp R., Wall E., Coffey M. *et al.* (2012) *J. Dairy Sci*. **95**:6103-6112.

Gilmour, A. R. G., Cullis, B.R. Welham, S.J. Thompson, R. (2006) ASReml User Guide Release 2.0. VSN International Ltd.

Haile-Mariam, M., Gonzalez-Recio O., and Pryce J. (2014) *J. Dairy Sci.* **97**:3173-3189.

Macdonald, K., Pryce J., Spelman R., Davis S., Wales W. *et al.* (2014) *J. Dairy Sci.* **97**:1427-1435.

Pryce, J. E., Gonzalez-Recio O., Thornhill J. B., Marett L. C., Wales W. J., *et al.* (2014). *J Dairy Sci* **97**:537-542.

Visscher, P., Bowman P., and Goddard M. (1994). *Livest. Prod. Sci.* **40**:123-137.

Williams, Y. J., Pryce J. E., Grainger C., Wales W. J., Linden N. *et al.* (2011). *J Dairy Sci* **94**:4715-4725.

Proc. Assoc. Advmt. Breed. Genet. 21: 29-32

BREEDING TO IMPROVE MEAT EATING QUALITY IN TERMINAL SIRE SHEEP BREEDS

A. Swan[1,4], T. Pleasants[2], and D. Pethick[3,4]

[1]Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351
[2]AgResearch, Ruakura, Hamilton, New Zealand
[3]School of Veterinary & Life Sciences, Murdoch University, Western Australia, 6150
[4]Australian Sheep Cooperative Research Centre, Armidale, NSW, 2351

SUMMARY

An economic value for sheep meat eating quality was derived using consumer taste panel sensory trait scores and willingness to pay data. Improving eating quality by one score generated a price premium to commercial producers of $0.15/kg relative to a carcass price of $4.50/kg. Eating quality was included in a breeding objective with growth and lean meat yield. Under selection index scenarios modelled, simultaneous improvement of all traits was only possible with genomic testing of male selection candidates due to antagonistic correlations involving yield, eating quality, intramuscular fat, and shear force. Economic gain could be increased by up to 20% compared to current industry selection indexes.

INTRODUCTION

Terminal sire breeders in Australia have made sustained genetic gains over a long period of time (e.g. Swan *et al.*, 2009). One of the ingredients for this success has been a simple breeding objective targeting increased growth and lean meat yield, which can be accurately evaluated from a young age using selection indexes based on body weight, and eye muscle and fat depth scanned on live animals. While the genetic gain in lean meat yield has contributed to lamb becoming a highly desirable product with increased consumer demand, care is needed to avoid making carcasses too lean and creating problems with eating quality.

The Sheep CRC has undertaken an extensive measurement program of carcass and eating quality traits on individual animals in the Information Nucleus (IN) flocks (Van der Werf *et al.*, 2010). In this study this data is used to develop an economic value for sheep meat eating quality which can be included in a breeding objective with economic traits including lean meat yield, and compare selection responses for indexes with and without eating quality and genomic selection.

MATERIALS AND METHODS

Eating quality traits. An eating quality trial based on consumer taste panels was conducted by the Sheep CRC on samples taken from IN animals born in 2009 and 2010. The design of the trial has been described in detail by Pannier *et al.* (2014), but briefly, ten samples were taken from both the loin and topside portions of carcasses of IN slaughter animals (n=1400+). These were prepared using a standard cooking method, and then consumed by the taste panels. The taste panel members scored each sample for five sensory eating quality traits on a 0 – 100 scale: odour, flavour, juiciness, tenderness, and overall liking.

The sheep meat industry uses the Meat Standards Australia (MSA) retail grading system with four effective grades (ungraded, and grades 3, 4, and 5). These are determined by an MSA score often derived as a linear function of the consumer eating quality traits (e.g. Johnston *et al.*, 2003) and expressed on the 0 – 100 scale. Genetic parameters estimated from the IN data (Mortimer *et al.*, 2015) show that within the loin and topside cuts genetic correlations between eating quality

* AGBU is a joint venture of NSW Dept. of Primary Industry and the University of New England

traits were always greater than 0.9, and the genetic correlation between loin and topside overall liking was 0.93. Therefore for the purposes of this study, topside overall liking is defined as the eating quality breeding objective trait (denoted as tmsa). After correcting for fixed effects including breed, the mean and standard deviation of tmsa in the IN data were 52.1 and 9.1 respectively. These values were used in the derivation of the economic value, defining the base distribution of eating quality in a commercial flock as $N(\mu = 52.1, \sigma = 9.1)$.

Two measured carcass traits, intramuscular fat (imf, %) and shear force (sf5, Newtons) are strongly related to eating quality, and these were also considered as selection criteria.

Economic value for eating quality. There were thresholds assumed on the tmsa trait scale which determine MSA retail grade, as shown in Table 1. Further, consumer willingness to pay surveys establish price relativities between retail grades, and the values assumed are also shown in Table 1. These two pieces of information can be used to derive an economic value for eating quality. Firstly, carcass value (CV) to the commercial producer can be expressed as:

$$CV = CWT \times [(1-m)p_N + mp_L \Sigma_i v_i r_i]$$

where CWT is carcass weight (kg), p_L is the price of lean meat for MSA grade 3 ($/kg), p_N is the price of the residual carcass component ("non-lean", $/kg), m is the ratio of lean meat yield, v_i is the probability of a carcass achieving MSA grade i from the base distribution of eating quality defined above, and r_i is the price relativity for MSA grade i as shown in the willingness to pay column of Table 1. Increasing tmsa by 1 score changes the MSA grade probabilities, increasing the probability of achieving a higher grade, and reducing the probability of ungraded meat. The carcass price premium associated with a 1 score increase in tmsa can be expressed as:

$$\Delta p_C = mp_L \Sigma_i (v_i^* - v_i) r_i$$

where v_i^* is the probability of achieving MSA grade i in the improved flock. The economic value for tmsa on a per carcass basis is now:

$$REV_{eq} = CWT \times \Delta p_C$$

Table 1. Lower threshold tmsa value for each MSA grade (Min tmsa), probabilty of MSA grade in base (v_i) and improved (v_i^*) flocks, and willingness to pay price relativities

MSA grade	Min tmsa	v_i	v_i^*	Willingness to pay
Ungraded	0	0.4079	0.3658	0.5
3	50	0.5140	0.5388	1.0
4	65	0.0722	0.0874	1.5
5	75	0.0059	0.0080	2.0

Breeding objectives including eating quality. A terminal sire breeding objective targeting a terminal sire x Merino dam commercial enterprise was developed using the SheepObject system (Swan *et al.*, 2007), based on growth (post-weaning body weight, pwt) and carcass traits including lean meat yield (lmy), dressing percentage (dress), carcass eye muscle depth (cemd), and carcass fat depth (cfat). The key price and production variables were carcass weight = 23kg, lean meat yield ratio = 0.56, and carcass price received by the producer = $4.50/kg. The eating quality economic value was added to the objective and converted to a per ewe joined basis by multiplying by a weaning rate of 0.95 lambs weaned per ewe joined, and discounted using a discount rate of 7%. This objective is denoted LMY_EQ. It was extended to include imf as a desired gains trait accounting for 5% of the total economic gain, and this objective is denoted LMY_EQ_IMF.

Genetic parameters for breeding objective and selection index traits. Genetic parameters were estimated using the IN animals in the eating quality trial. To increase confidence in the consistency of correlations, a multivariate analysis was performed using the R package MCMCglmm (Hadfield, 2010) simultaneously including the breeding objective traits pwt, lmy,

dress, cemd, cfat, and tmsa, and potential selection index traits weaning weight (wwt), post-weaning eye muscle (pemd) and fat depth (pfat), imf, and sf5.

Prediction of genetic gains. Genetic gains from index selection over 10 years were calculated for the LMY_EQ and LMY_EQ_IMF objectives and compared to gains from the current Carcass+ (CPLUS) industry objective. Gains were calculated for a terminal sire breeding flock of 300 ewes with 10 sires mated annually and a weaning rate of 1.3 lambs per ewe joined. Selection intensities were 2.328 for males and 0.860 for females and generation intervals were 2.6 for males and 3.2 for females. These figures were derived from the LAMBPLAN genetic evaluation database. For each objective response was calculated from two scenarios. In the first, phenotypes were available on the base traits of wwt, pwt, pemd, and pfat. The second added genomic predictions on young males for base traits (all objectives) and the carcass traits lmy, cemd, cfat, dress, imf and sf5 (LMY_EQ and LMY_EQ_IMF only) using the accuracies for genomic predictions currently used in the LAMBPLAN genetic evaluation system (Swan *et al.*, 2014).

RESULTS AND DISCUSSION

Economic values for the LMY_EQ and LMY_EQ_IMF breeding objectives are shown in Table 2. The economic value for eating quality was $3.21 per ewe joined. Relative to a carcass price of $4.50 per kg, the price premium for a 1 score increase in eating quality was $0.15 per kg. This premium is currently not realised by commercial producers as there is no supply chain feedback for eating quality at the level of individual carcasses.

Genetic correlations for lmy and tmsa with other economic and selection criteria traits are also shown in Table 2. There was a small antagonistic correlation between lmy and tmsa (-0.12), and large antagonisms involving lmy, tmsa, imf and sf5. The latter two traits are important selection criteria for eating quality, with improved eating quality associated with higher imf (0.31) and lower sf5 (-0.31). However, the reverse is true for lmy, which is strongly associated with lower imf (-0.55) and higher sf5 (0.40). These antagonisms limit the genetic gain which can simultaneously be made in lean meat yield and eating quality.

Mortimer *et al.* (2015) estimated genetic correlations for a wider range of sensory scores and found that eating quality in the loin was more strongly associated with imf, while for the topside, shear force had the stronger association. We note that it is possible to extend the approach outlined above to calculate economic values separately for different carcass cuts.

Table 2: Economic values for LMY_EQ and LMY_EQ_IMF breeding objectives ($/ewe joined), and genetic correlations used in index predictions for lmy (r_g lmy) and tmsa (r_g tmsa)

Trait	Units	LMY_EQ	LMY_EQ_IMF	r_g lmy	r_g tmsa
pwt	kg	1.834	1.834	0.10	0.03
lmy	%	1.879	1.879	1.00	-0.12
dress	%	2.042	2.042	0.00	-0.10
cemd	mm	3.267	3.267	0.10	-0.17
cfat	mm	-0.966	-0.966	-0.66	-0.05
tmsa	0 – 100	3.211	3.211	-0.12	1.00
imf	%		15.727/7.867[A]	-0.55	0.31
sf5	Newtons			0.40	-0.31

[A]Index dependent (15.727 with base traits measured, 7.867 with genomic predictions added)

Genetic gains in Table 3 show that the current industry objective CPLUS is predicted to produce significant gains in growth rate (pwt), lean meat yield and carcass eye muscle, at the expense of a reduction in eating quality. The base and genomic testing scenarios produce very similar outcomes for CPLUS. For the LMY_EQ objective, gains in growth and lean meat yield were further enhanced compared to CPLUS and while eating quality gain was still negative, it was

closer to zero. Genomic testing increased economic gain for this index by 9%, and compared to the CPLUS scenarios by up to 20%. When only base traits were available, including the restriction on imf in LMY_EQ_IMF resulted in a negative response in lean meat yield, a strong positive response in carcass fat, and a small positive response in eating quality. Compared to the economically optimum index, economic gain was reduced by 29%. By adding genomic testing in LMY_EQ_IMF favourable responses were achieved in lean meat yield, eating quality, imf and sf5, and economic gain was reduced by only 9%.

Table 3: Trait gains over 10 years for CPLUS, LMY_EQ and LMY_EQ_IMF objectives with base and genomic selection criteria, total dollar gain and economic efficiency

Trait	Units	CPLUS		LMY_EQ		LMY_EQ_IMF	
		base	genomic	base	genomic	base	genomic
pwt	kg	4.28	4.46	5.36	5.30	5.41	4.98
lmy	%	1.07	1.10	1.43	1.46	-0.50	0.19
dress	%	1.19	1.27	0.63	0.80	0.72	0.81
cemd	mm	1.55	1.66	0.71	0.85	0.39	0.62
cfat	mm	0.24	0.27	-0.25	-0.21	0.48	0.29
tmsa	0 – 100	-1.25	-1.34	-0.44	-0.19	0.10	0.64
imf	%	-0.26	-0.27	-0.34	-0.28	0.04	0.10
sf5	Newtons	1.14	1.17	1.88	1.16	0.55	-0.68
$ gain		13.11	13.66	14.93	16.47	11.61	14.95
Efficiency		80	83	91	100	71	91

CONCLUSIONS

The breeding objectives presented in this study demonstrate that terminal sire breeders can simultaneously improve growth, meat yield and eating quality, albeit with restrictions due to antagonistic genetic correlations between traits. To realise the benefits of the breeding objectives it is necessary to increase the accuracy of genetic evaluations of carcass traits and eating quality traits including intra muscular fat and shear force. Genomic testing is one way to achieve this increased accuracy.

ACKNOWLEDGEMENTS

This work was partly funded by MLA Project B.SGN.0127.

REFERENCES

Hadfield, J., D. (2010) *J. Stat. Softw.* 33:1–22.

Johnston, D. J., A. Reverter, D. M. Ferguson, J. M. Thompson, and H. M. Burrow. (2003) *Aust. J. Agric. Res.* **54**:135.

Mortimer, S. I., A. A. Swan, L. Pannier, A. J. Ball, R. H. Jacob, and J. H. J. Van der Werf. (2015) In: *Proc. Assoc. Advmt. Anim. Breed. Genet.* Vol. 21.

Pannier, L., G. E. Gardner, K. L. Pearce, M. McDonagh, A. J. Ball, R. H. Jacob, and D. W. Pethick. (2014) *Meat Sci.* **96**:1076.

Swan, A. A., D. J. Brown, and R. G. Banks. (2009) In: *Proc. Conf. Assoc. Anim. Breed. Genet.* Vol. 18. p. 326.

Swan, A.A., D. J. Brown, H.D. Daetwyler, B.J. Hayes, M.J. Kelly, N.Moghaddar, and J.H.J. Van der Werf. (2014) In: *Proc. 10th World Congr. Genet. Appl. Livest. Prod.* p. 334.

Swan, A. A., J. H. J. van der Werf, and K. D. Atkins. (2007) In: *Proc. Conf. Assoc. Anim. Breed. Genet.* Vol. 17. p. 483.

Van der Werf, J., B. Kinghorn, and R. Banks. (2010) *Anim. Prod. Sci.* **50**:998.

Proc. Assoc. Advmt. Breed. Genet. 21: 33-36

ACCOUNTING FOR THE COST OF REPRODUCTIVE TECHNOLOGIES DURING SELECTION IN SHEEP BREEDING PROGRAMS

T. Granleese[1,2], S.A. Clark[1,2], A.A. Swan[2,3], J.H.J. van der Werf[1,2]

[1] School of Environmental and Rural Science, University of New England, Armidale, Australia, 2351
[2] Cooperative Research Centre for Sheep Industry Innovation, Armidale, 2351
[3] Animal Genetics and Breeding Unit, Armidale, 2351

SUMMARY

Female reproductive technologies, such as MOET and JIVET, have been shown to increase the rate of genetic gain. However, they incur substantial costs to breeders using them. In this work, optimal contribution selection was used to find the balance between genetic merit, co-ancestry and cost of reproductive technologies in sheep breeding programs. To offset the cost of using the reproductive technologies, breeders received a premium based on the value of the genetic gain achieved by the ram buyers. Australian terminal sire and Merino breeding programs were simulated, using industry indexes. For the terminal sire breeding program, the premium needed to be greater than 50% before reproductive technologies were used. In the Merino breeding program, where the standard deviation of the index is 3 times higher than the terminal index, reproductive technologies were used with lower premiums (6% and 32% premiums, respectively). For both breeding programs, the rate of genetic gain increased with more allocations of reproductive technologies. There was also a higher proportion of JIVET assigned compared to MOET, due to a lower cost per lamb. The benefits of genomic selection were greatest in the merino program, due to the higher use of JIVET. Assigning costs of reproductive technologies allows for robust and practical breeding programs to be designed.

INTRODUCTION

Female reproductive technologies such as multiple ovulation and embryo transfer (MOET) and juvenile *in vitro* fertilisation and embryo transfer (JIVET) have been shown to increase rates of genetic gain by increasing selection intensity and decreasing generation interval. However, use of these reproductive technologies can result in higher rates of inbreeding through selection emphasis on elite families rather than elite individuals within families. JIVET has become a more viable reproductive technology option since the introduction of genomic testing as it increases the selection accuracy of juvenile animals. Optimal contribution selection (Wray and Goddard 1994) is an effective selection tool to maximise genetic gain while maintaining sustainable rates of inbreeding. It has been shown to be an effective method in optimising allocation of reproductive technologies in stochastic simulations where long-term genetic merit and co-ancestry are balanced, and could also be used to account for the cost of reproductive technologies. To warrant breeders investing in expensive genetic acceleration programs using reproductive technologies, a premium per ram should be paid by ram buyer which reflects the proportion of total benefits of increased genetic gain that flows to the buyer.

This paper aims to explore, via a stochastic simulation study, the optimal allocation of reproductive technologies using optimal contribution selection and accounting for the cost of reproductive technologies.

MATERIALS AND METHODS

We evaluated reproductive technologies in both terminal sire and Merino breeding programs, with selection based on industry indexes ("Lamb 2020" for terminal sires and "Merino Production" for Merinos, from http://www.sheepgenetics.org.au). A closed breeding nucleus of 250 sheep were stochastically simulated and then bred over 15 years over 100 replications. For each scenario an unrelated base population and subsequent overlapping generations were generated. True breeding values and phenotypes were simulated with variances and co-variances used in the Sheep Genetics evaluation system (Huisman et al. 2008). Each year individual animals had breeding values estimated (EBV) via pedigree based on multi-trait Best Linear Unbiased Prediction (BLUP) using ASReml software (Gilmour et al. 2009).

There were up to three types of matings allowed in the breeding programs: 1) artificial insemination or natural mating (AI/N) 2) MOET and 3) JIVET. Individuals were selected for one of these mating types using optimal contribution selection (Wray and Goddard 1994). The objective function M+C-R was optimised, where $M=x'b$ where x is a vector with genetic contributions and b is a vector with BLUP EBV on n selection candidates; $C=\lambda \cdot x'Ax$ where λ is a negative value to maintain inbreeding rate at 1% (±0.05) per generation, and A is the pedigree relationship matrix among selection candidates; $R = \beta \cdot x'd_t$ where d_t is cost per lamb resultant from reproductive technology (t) (Table 4). β is a scaling factor to express the cost of using reproductive technologies on the same scale as the additional genetic merit of the animals selected. Hence, $\beta= W_2/(p \cdot W_1)$ where W_2 is the number of progeny born in the nucleus each year multiplied by 2, W_1 is the number of commercial animals bred by rams from nucleus multiplied by the cumulative discount expression (CDE) of genetic superiority (Hill 1974) and p is a premium paid by ram buyers which is the equivalent to a proportion of the genetic benefit (resultant change in index) they will receive in their commercial flock(s) of 5000 ewes and paid back to the ram breeder. The premiums paid (p) varied at levels of 0.06, 0.32 and 0.64 as representations of low, medium and high premiums. The cost of reproductive technology and average number of lambs born per technology are shown in Table 1. Costs included drugs, professional services and price of failed transfers. The fecundity was averaged from previous studies and costs averaged from questionnaires completed by advanced reproduction companies in Australia.

For each breeding program and index the impact of genomic selection (GS), assuming all animals had genomic information available at birth, was assessed. The cost of GS was not accounted for in this study. Genomic information was modeled following the method of Dekkers (2007) which simulates a genomic breeding value as a correlated trait with a heritability of 0.999 and a correlation r to the measured trait, where r is the accuracy of the genomic breeding value for each trait. The accuracy of the genomic test varied for each trait (Swan *et al.* 2014).

Table 1. **Average cost per lamb and number of progeny per program.**

Technology	Cost ($/lamb)	Ave. progeny (n)
No mating	0	0
AI/N	20	1
MOET	160	4
JIVET	130	8

RESULTS AND DISCUSSION

In the terminal sire breeding programs, the proportion of benefit paid to the breeder had to reach 0.64 before any reproductive technologies were assigned. At a premium of 0.64 for the terminal breeding program, it was observed an increase in annual response in the nucleus of 5% was associated with 3-5% of lambs born via JIVET technology (Table 2). By contrast, the Merino breeding programs had reproductive technologies assigned when the proportion of benefit was as low as 0.06 (Table 2). This is expected with the Merino index having a larger index dollar genetic standard deviation and therefore a higher value of the genetic gain achieved. In the Merino index it was observed that as the premium increased, the allocation of reproductive technologies increased, as did the genetic gain (Table 2). In the AI/N+ MOET+JIVET program (using GS) a 75% higher rate of genetic gain was observed for the 0.64 premium scenario compared to the 0.06 premium scenario.

Table 2. Proportion of lambs born to reproductive technologies, number of dams required to breed 250 lambs, annual genetic gain (G/yr) and average generation interval (L) in respective breeding programs using terminal sire Lamb 2020 and Merino MP indexes at 1% increase in inbreeding per generation.

Proportion of benefit paid to breeder	AI	MOET	JIVET	Dams Used	G/yr ($)	L
Lamb 2020						
AI/N + MOET + JIVET (GS)						
0.06	1.00	0.00	0.00	271	1.24	1.83
0.32	1.00	0.00	0.00	268	1.28	1.86
0.64	0.95	0.00	0.05	259	1.31	1.81
AI/N + MOET + JIVET						
0.06	1.00	0.00	0.00	269	1.13	1.94
0.32	1.00	0.00	0.00	273	1.18	1.98
0.64	0.97	0.00	0.03	268	1.19	1.91
MP						
AI/N + MOET + JIVET (GS)						
0.06	0.95	0.00	0.05	261	2.26	1.87
0.32	0.77	0.04	0.19	221	2.82	1.46
0.64	0.36	0.10	0.54	136	3.96	1.21
AI/N + MOET + JIVET						
0.06	0.94	0.01	0.05	268	1.32	1.98
0.32	0.82	0.03	0.15	233	1.85	1.51
0.64	0.41	0.12	0.47	129	2.02	1.38

SEM for dams used ≤±4.3, ΔG/yr ≤±$0.05, L ≤± 0.05 for all breeding programs.

The impact of using genomic selection in breeding programs varied between indexes. For terminal breeding programs, when comparing the same breeding program with and without genomic selection, a 6-10% increase in annual response was found (Table 2). This low increase is expected with the key traits measured within 6 months of life and very little JIVET assigned, where genomic selection would be most beneficial in a terminal breeding program. However, we observed increases of up to 96% in the Merino breeding program. The larger response with genomic selection is expected with all traits measured after one year of age and key traits, such as number lambs weaned, not phenotypically measured.

In both Merino and Terminal breeding programs where only up to 5% of progeny resulting from JIVET, slight increases in annual genetic gain of up to 7% were found (Table 2). It is expected that these matings are performed on females that are outliers in the population and would produce progeny that significantly contribute to subsequent generations (i.e. ram progeny selected in future years). These matings are strategic and may be observed in current breeding practices in industry.

When the cost of reproductive technology is accounted for during selection and the premium paid by buyers is zero, the extra income received through higher performance (i.e. increased fleece weight, etc.) facilitated via genetic gain is not high enough to justify its use. Therefore, ram buyers who want genetically superior rams derived from the use of advanced reproductive technologies, will need to pay some form of premium to the ram breeders. However, past experience has shown that the value of premiums paid by ram buyers can be somewhat arbitrary and usually follow market trends rather than benefit captured by the buyers (Banks *et al.* 2014).

CONCLUSIONS

Applying a true cost to reproductive technologies during the optimal contribution selection method delivered practical mating solutions in breeding programs. A premium paid as a proportion of the benefit received by ram buyers for stud rams provides an avenue to justify and recover the costs of using reproductive technologies by stud breeders. Higher premiums paid resulted in more reproductive technologies used and as consequence, faster annual rates of genetic gain. Genomic selection facilitated better selection decisions on younger selection candidates and provided the most benefit in the Merino breeding programs, where most traits in the index are measured later in life, not measured at all, or are hard to measure. A terminal sire program using the Lamb 2020 index had limited justification for investment in reproductive technologies to accelerate genetic gain due to comparatively low rates of true dollar genetic gain.

REFERENCES

Banks R.G., Brown D.J., Swan A.A. (2014) Animal Breeding and Genetics Unit: *Response to AWI Industry Consultation Document*.Available from: http://www.wool.com/

Dekkers J.C.M. (2007) *J. Anim. Breed. Genet.* **124**:331.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., Thompson R. (2009) ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP11ES, UK.

Hill W.G (1974) *J. Anim. Breed. Genet.* **124**:331.

Huisman A.E., Brown D.J, (2008) *Aust. J. Exp. Agr.* **48**:1186.

Swan A.A., Brown D.J., Daetwyler H.D., Hayes B.J., Kelly M., Moghaddar N., van der Werf J.H.J. (2014) *Proc. 10th World Congr. Gen. Appl. Livest. Prod., Vancouver, Canada.*

Wray N.R. and Goddard M.E. (1994) *Genet. Sel. Evol.* **26**:431.

Proc. Assoc. Advmt. Breed. Genet. 21: 37-40

GENETIC EXPRESSIONS AND ECONOMIC IMPACT OF PERFORMANCE RECORDING IN MULTI-TIERED SHEEP BREEDING SCHEMES

B.F.S. Santos[1, 2], T.J. Byrne[1], B. Visser[1], J.H.J. van der Werf[2], J.P. Gibson[2] and P.R. Amer[1]

[1]AbacusBio Limited, P O Box 5585, Dunedin 9058, New Zealand

[2]School of Environmental & Rural Science, University of New England, Armidale, NSW

SUMMARY

This research initiative explores the value of performance recording in the multiplier flock of multi-tiered sheep breeding schemes integrating commercial and breeding structures. Discounted gene flow theory was used to model the flow of genes from the selected multiplier rams that are mated to commercial ewes. A model which predicts genetic trends and gene flows at an individual trait level was combined with trait economic values to aggregate the total industry benefits associated with alternate recording and selection strategies. While recording efforts and molecular technology adds significant costs to the breeding schemes, their application in multiplier flocks also adds considerable margins by increasing accuracy of selection with consequent higher rates of progress.

INTRODUCTION

In New Zealand, the sheep breeding sector is part of a wider multi-tier structure. The majority of sheep breeding is structured in two tiers, nucleus and commercial, with the special case where an intermediate tier multiplies rams to be mated in commercial flocks; a multiplier tier. The highly varied production systems across the country require that breeders, and ultimately commercial farmers, have the ability to rank selection candidates on combinations of performance traits applicable to their systems. Recently, an increasing number of sheep breeder conglomerates and groups have been formed allowing larger investments in breeding program design. It is important to increase the accuracy of prediction, and consequently the rate of genetic progress, which in lower tiers is equal to the progress obtained in the nucleus, although more accurate selection in the multiplier will reduce the genetic lag that exists between nucleus and commercial flocks. Given the opportunity to integrate information flow across tiers through the application of new molecular and phenotype data collection technologies, higher genetic gains can also be achieved by expanding performance recording in the commercial flock. The role of multiplier flocks has increased due to demand for more productive and affordable rams which originate from flocks that apply up to date selection technology. In multi-tiered breeding schemes the informed selection of replacement ewes and rams for use in lower tiers results in added value to the commercial farmer. This research examines breeding scheme design and the integration of breeding and commercial performance data to achieve faster genetic gain. This study investigates the value of selection based on performance recording in the multiplier tier of a breeding scheme in which commercial sheep production is part of a vertically-integrated breeding structure.

MATERIAL AND METHODS

This research was based on multiple trait deterministic simulation using New Zealand industry parameters. The primary focus was on meat sheep production for a commercial operation where there is limited use of terminal sires, which means that flocks in this operation are mainly using livestock selected for maternal traits. A number of scenarios based on different strategies for implementing recording in the multiplier flock were simulated and their benefits evaluated.

Simulation scenarios. The simulation scenarios accounted for the recording strategy, the use of genomic selection in the different tiers, selection and culling of candidates based on breeding index (full trait range) or production index (limited trait range), and alternative replacement

policies. The policies for replacing ewes in the nucleus flock assume candidates are available from two sources; young female hoggets from the nucleus flock, or older mixed-age proven ewes from the multiplier flock. Full recording practices and parentage were assumed within the nucleus flock, whilst simple trait performance recording, such as pregnancy scanning, live weight and body condition score, were assumed in the multiplier flock. The simulation scenarios modelled in this study were: *Base scenario*, no performance recording in the multiplier tier; *Scenario 1*, performance recording in the multiplier tier implemented in Year 1, including DNA parentage, the top 5% of rams tested on 50K SNP chip, the top 15% of rams tested on 5K SNP chip (Genomic Selection, GS), selection on a breeding index, and young female hoggets from the nucleus flock selected as replacements; *Scenario 2*, same recording as Scenario 1, but without GS; *Scenario 3*, performance recording in multiplier tier from Year 1 without parentage, and selection based on phenotypic performance of ewes and a phenotypic selection index; *Scenario 4*, same as Scenario 1, but with selection of mixed-age proven ewes to replace the nucleus flock; *Scenario 5*, same as Scenario 2, but with selection of mixed-age proven ewes to replace the nucleus flock; and *Scenario 6*, where GS is implemented based on genotyping of the flock, but without phenotypic recording.

Genetic trends. Average genetic trends were calculated for each trait in each tier based on deterministic prediction, for which calculations included trait specific genetic parameters and accuracies, selection intensity and generation interval. The calculation of genetic progress in ewes born in the nucleus before the base year of selection in the multiplier flock (Year 1) was assumed as a linear relationship between age at selection and the annual rate of genetic progress for the different traits. The average genetic merit of the flock was calculated based on recursive equations which account for the proportion of animals in each age class. The selection differential and consequent genetic progress was calculated using selection index theory principles combined with the methodology proposed by Ducrocq and Quaas (1988). The average genetic merit in the different tiers was calculated based on the reported response to selection values obtained from a selection index model from van der Werf (1999). These responses were converted into selection differentials which form the basis to calculate genetic trends across tiers and scenarios. The progeny merit in a given year for a specific trait is equivalent to the mean genetic merit of sires and dams selected to produce the progeny, calculated as described above, and the selection differential applied in the progeny. Accuracies of GS predictions were modelled as additional traits, genetically and phenotypically correlated with traits included in the selection index model (van der Werf, 2009).

Discounted genetic expressions. The discounted genetic expression coefficients (DGEs) account for the proportion of superiority transmitted over time to an individual's descendants through transfer of genes. The expressions of genes in different age classes, discounted at 7% per annum, were calculated through a series of transition matrices for a self-replacing ewe (Amer, 1999). In the current study, genetic expressions were used to model the flow of genes from multiplier rams once they pass into the commercial flock for mating. Discounted genetic expression coefficients for traits expressed at different time points were grouped as vectors (with dimension equal to the year of expression) of: annually expressed traits, traits expressed at end of life or cull traits in ewes, traits expressed at birth, and at slaughter in lambs. Age composition, and reproductive and survival rates used in the calculations of DGEs within the gene-flow model were obtained from industry data.

Cost-benefit analysis. Benefits were calculated across all tiers for the different scenarios. The genetic merit of cohorts combined with DGEs and trait economic values were used to generate the monetary impact observed in the commercial tier with and without recording in the multiplier flock for all scenarios. The total economic impact is presented as the additional benefits of the implementation of varying recording practices in the multiplier tier, relative to base scenario, after

accounting for the costs of parentage testing, genomic selection, electronic identification, recording, and genetic evaluation. The economic impact was estimated for a large commercial operation with 180,000 ewes lambing per year.

RESULTS AND DISCUSSION

Within scenarios, there were significant differences in traits' selection differentials between tiers. Across scenarios, selection differentials were higher in the nucleus in comparison to the multiplier and commercial flocks, and they differed considerably between scenarios. The rates of genetic progress in the commercial tier of scenarios 1 and 4 were the highest, followed by scenarios 2, 5 and 6 which had similar rates. The rates of progress for scenarios 1 and 4, and for scenarios 2 and 5 were nearly identical. The lowest progress was achieved in Scenario 3. In time, genetic progress tended to stabilize and become constant across tiers, although differences in average merit (genetic lag) between tiers and between scenarios remained. These results also reflect the variation in the timing of actual expressions of traits; e.g. lamb birth trait impacts occur first, and cull ewe trait impacts occur last. Table 1 presents, for each trait, the genetic lag between the nucleus and the commercial flock in the different scenarios. A reduction of approximately two years in the lag between the nucleus and the commercial flock was achieved when recording was implemented in the multiplier tier for all scenarios, except for Scenario 3.

Table 1. Genetic lag (years) in year 20 between the nucleus and the commercial tier in different scenarios, where Com_c represents the commercial flock after recording is implemented in the multiplier.

	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6	
Trait (Abbrev.)	Com	Com_c	Com	Com_c	Com	Com_c	Com	Com_c	Com	Com_c	Com	Com_c
Carcase weight	7.92	5.78	7.93	5.90	8.29	7.01	7.93	5.71	7.93	5.80	7.86	5.53
Weaning weight	7.92	5.78	7.93	5.89	8.36	7.48	7.93	5.72	7.93	5.82	7.86	5.54
Number of lambs born	8.79	6.63	8.79	6.59	8.99	6.79	8.78	6.65	8.77	6.93	8.74	6.39
Ewe mature weight	8.79	6.77	8.79	7.14	8.97	6.56	8.80	6.50	8.82	6.75	8.53	5.68
Ewe BCS	8.78	6.04	8.78	5.92	9.22	9.22	8.76	6.58	8.77	6.34	9.21	7.94
Survival mat	8.78	6.27	8.78	6.58	9.44	9.44	8.78	6.27	8.78	6.58	9.68	9.68
Weaning weight mat	8.79	6.65	8.79	6.59	9.44	9.44	8.79	6.59	8.77	7.00	8.65	6.09
Stayability	8.44	5.84	8.43	4.73	8.98	8.98	8.43	6.23	8.42	5.24	8.87	7.74
Lamb survival	8.00	5.80	11.1	11.1	11.1	11.1	8.00	5.80	11.1	11.1	8.00	5.80

Figure 2 presents the difference from the Base scenario of the cumulative net present value (NPV), estimated as the sum over time of the present values of benefits and costs discounted at 7% per annum. The NPV for scenarios 4 and 5 lie exactly underneath those for scenarios 1 and 2 respectively, as a consequence of nearly identical rates of genetic progress. Scenarios assuming parentage assignment (i.e. all scenarios but Scenario 3) presented an early spike in costs associated with DNA testing in the multiplier flock which stabilised over time. Yearly recording costs also include trait measurement, electronic identification (EID) and genetic evaluations. The marginal (relative to Base scenario) commercial flock benefit, which includes the multiplier flock as a significant component of the commercial operation, accumulated steadily after a delay and then stabilized after ten years as a constant flow of benefits. Benefits stabilized after a given number of years because genetic progress becomes constant. Losses are reversed after 5 – 7 years in all scenarios with exception of Scenario 3, and in the long run, annual benefits exceed annual costs by a factor of two. Scenario 3, which assumed phenotypic selection without parentage assignment, did not result in enough extra genetic progress to offset the performance recording costs.

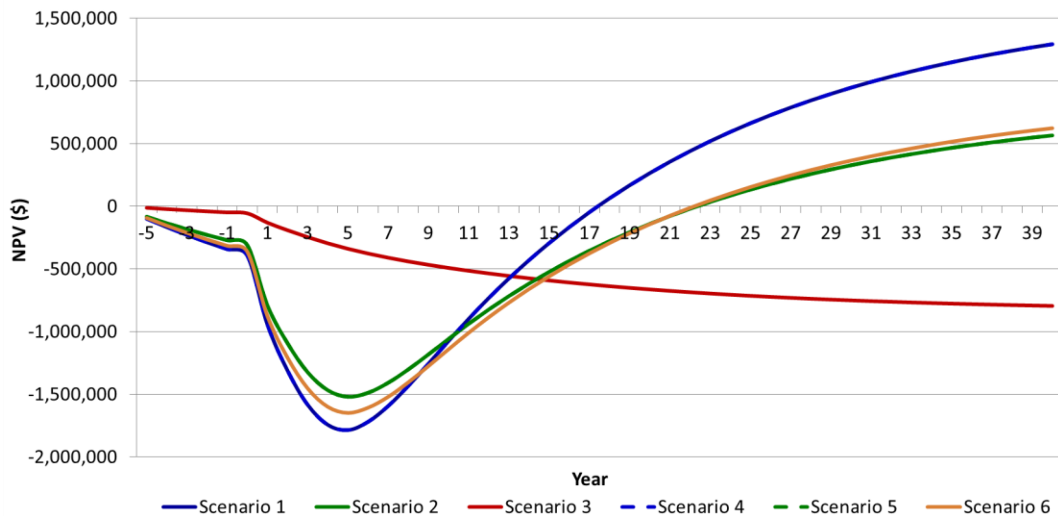


Figure 2- Cumulative net present value across scenarios assuming different strategies of performance recording in the multiplier tier of multi-tiered breeding scheme.

Selection intensity and age structure differences between scenarios assuming mixed-age ewe replacements did not result in changes in NPV when compared to the policy where young ewe lambs are selected as flock replacements. Results demonstrated that higher genetic progress can still be achieved through more accurate selection based on selection of proven mixed-age ewes.

The results demonstrated the potential that GS has to increase genetic progress in the breeding scheme. Increased progeny merit and reduced genetic lags were achieved in scenarios 1 and 4, which assume GS, when compared to equivalent scenarios without GS. The implementation of a GS strategy was cost-effective when including current costs of genetic testing. The strategy where GS without phenotypic recording is undertaken was promising, and its feasibility relies on the accuracy of genomic predictions.

CONCLUSION

The results of this study demonstrate that performance recording in the multiplier tier can reduce the long genetic lag between the nucleus and commercial flocks in multi-tiered breeding programs. The results also demonstrate that economic benefits can be generated by implementing recording in the multiplier tier. Such recording is justified if the breeding scheme captures all of the benefits through value added to slaughter lambs and replacement females produced in the commercial tier. The investment in genotyping was the major expense in the scenarios where it was applied. Strategies where proven mixed-age ewes are selected as replacements of the nucleus flock produce identical margins to those which select young ewes, offsetting eventual increases in generation interval. It is important to note that the biggest benefits came from a combination of full recording of functional traits plus parentage assignment.

REFERENCES

Amer, P. R. (1999). *New Zealand J. Agri.Res.*, **42**(3), 325–336.

Ducrocq, V. & Quaas, R.L. 1988. *J. Dairy Sci*, **71**: 2543-2553.

van der Werf, J.H.J. (1999). Multi-trait selection index MTINDEX20T model. http://www-personal.une.edu.au/~jvanderw/software.htm

Proc. Assoc. Advmt. Breed. Genet. 21: 41-44

WHY CAN WE IMPUTE SOME RARE SEQUENCE VARIANTS AND NOT OTHERS?

B.J. Hayes[1,2,3], P.J. Bowman[1,3], H.D. Daetwyler[1,2,3] and M.E. Goddard[1,3,4]

[1]AgriBio, Centre for AgriBioscience, Biosciences Research, DEDJTR, Victoria, Australia
[2]Biosciences Research Centre, La Trobe University, Victoria, Australia
[3]Dairy Futures Cooperative Research Centre, Victoria, Australia
[4]Faculty of Veterinary & Agricultural Sciences, University of Melbourne, Victoria, Australia

SUMMARY

We investigated how well rare variants can be imputed, using 1000 bull genomes sequence data set (1147 sequences) as a reference for imputation, and a target set of dairy cattle with 630K SNP genotypes, that were also genotyped for four rare recessive defects (BLAD, CVM, HH1 and JH1). The proportion of carriers correctly imputed ranged from 1, for JH1, to 0.04 for CVM. There was a general trend for the proportion of carriers correctly imputed to increase as the frequency of the rare allele increased. CVM did not follow this trend – the frequency of the rare allele for this locus was 10 times higher than for BLAD, but proportion of carriers correctly imputed was much lower than BLAD. On closer inspection, the core haplotype of sequence variants common to all CVM carriers was found in many non-carriers, and even in breeds other than Holstein (the disease has only been reported in Holstein). This was in contrast to JH1, where the core haplotype shared by carriers was unique to carriers, and was not found in other breeds. These results shed light on why we can impute some rare sequence variants well, while others are very difficult to impute.

INTRODUCTION

One motivation for using whole genome sequence data in genomic prediction and genome wide association studies (GWAS) is that whole genome sequence data will include rare variants which may explain some variation in the targeted complex traits. SNP arrays have limited power to capture this variation, as the SNP on these arrays are selected to have high minor allele frequency (MAF), and are therefore unlikely to be in high linkage disequilibrium with the rare variants. The cost of whole genome sequencing is currently too high to sequence the large number of individuals required for accurate genomic predictions or powerful GWAS. Therefore an alternative strategy has been proposed – sequence a proportion of the individuals in the population (1000 Genomes Project Consortium *et al*. 2012), or preferably the key ancestors of the population (eg Daetwyler *et al*. 2014), and then impute the sequence variants into all individuals genotyped with SNP arrays. How much variation is explained by rare variants in subsequent genomic predictions or GWAS will then depend on how much the rare variants truly explain, and the accuracy of imputing these rare variants.

Here we investigate how well rare variants can be imputed, using 1000 bull genomes sequence data set as a reference, and a target set of dairy cattle that were actually genotyped for four rare recessive defects. In order to gain insights into parameters affecting accuracy of imputation of rare variants, we investigated the length of core haplotype surrounding the disease allele for each recessive defect, the occurrence of this haplotype (minus the disease allele) in non-carriers and the frequency of this haplotype in breeds other than the one in which the disease occurs.

MATERIALS AND METHODS

Carrier status (from genotyping the causal mutation was available for four recessive diseases - Bovine leukocyte adhesion deficiency (BLAD, Shuster *et al*. 1992), complex vertebral malformation (CVM, Thomsen *et al*. 2006), Holstein Haplotype 1 (HH1, Adams *et al*. 2012) and

Jersey Haplotype 1 (JH1, Sonstegard *et al*. 2013). Genotypes for these mutations were available for 5987 Holstein (BLAD, CVM), 707 Holstein (HH1) and 16 Jersey bulls (JH1), respectively, as well 630K Bovine HD real or imputed SNP genotypes (eg. Erbe *et al*. 2012). In order to impute the BLAD, CVM, HH1 or JH1 genotypes into these animals, to compare with their actual genotypes, we used a reference data set of 1147 bulls and cows of 20 breeds with whole genome sequence, These reference animals were sequenced at between 4 and 40 times coverage, with an average of 11.2x, from 1000 bull genomes Run4.0. The breeds with largest number of sequenced individuals were Holstein, Angus and Fleckvieh. Variant calling and filtering was as described by Daetwyler *et al*. (2014). Variants with less than 4 copies of the minor allele were removed. We checked that all known carriers of BLAD, CVM, HH1 or JH1 that had whole genome sequence data (eg were part of the 1000 bull genomes) were genotyped correctly for these mutations, this was the case. Two imputation strategies to impute sequence variants into the target populations were tested, Fimpute (Sargolzaei *et al*. 2014) or Beagle phasing followed by Minimac imputation (Howie *et al*. 2012). Differences between these programs are that Fimpute uses full pedigree information, while Minimac does not, and Fimpute considers variable length haplotypes, starting from long haplotypes, when deciding if a pair of animals share a haplotype. Actual genotypes of the recessive lethals for target animals were not included when target animals were imputed to whole genome sequence genotypes. Imputed genotypes were then compared to actual genotypes for these defects.

RESULTS AND DISCUSSION

The proportion of genotypes imputed correctly was close to one for all loci, Table 1.

Table 1. Proportion of genotypes and proportion of carriers correctly imputed for four genetic defects.

	BLAD	CVM	HH1	JH1
Chromosome	1	3	5	15
Location (bp)	145114963	43412427	63150400	15707169
Frequency	0.001	0.010	0.025	0.156
Bulls genotyped in target population	5987	5987	707	16
Genotypes imputed correctly				
Fimpute	5970	5836	701	16
Minimac	5860	5860	705	16
Prop. genotypes imputed correctly				
Fimpute	0.997	0.97	0.99	1.00
Minimac	0.98	0.98	0.997	1.00
Number of carriers	17	123	35	5
Carriers correctly imputed				
Fimpute	13	5	29	5
Minimac	11	12	33	5
Prop. carriers correctly imputed				
Fimpute	0.77	0.04	0.83	1.00
Minimac	0.65	0.10	0.94	1.00

However this is a poor measure of how well imputation has performed for rare variants, given the high probability of filling in the correct genotype by chance (a very high proportion of animals are homozygous for the non-disease allele).

A better measure of how well imputation has performed is the proportion of carriers correctly imputed – for GWAS and genomic prediction, this will determine how well the SNP effect can be estimated. This ranged from 1, for JH1, to 0.04 for CVM. There was a general trend for the proportion of carriers correctly imputed to increase as the frequency of the rare allele increased. The imputation of CVM genotypes did not follow this trend – the frequency of the rare allele for this locus was 10 times higher than for BLAD, but the proportion of carriers correctly imputed was much lower than for BLAD.

To investigate why this might be the case, and given imputation is based on haplotype information shared between individuals, we determined the length of haplotype in the sequenced bulls (from the 1000 bull genomes project) surrounding the rare allele of each locus that was common between all carriers, the "core haplotype". To do this, we allowed for sequencing error, such that the shared haplotype was considered to end only when there were at least two differences in the alleles of the haplotype of the carriers (eg one difference was considered to be likely sequencing error - in fact there were only one or at most two instances of this per disease). HH1 had the longest core haplotype, while CVM had the shortest, Table 2. We then investigated how many non-carriers amongst all the Holstein sequenced bulls (for BLAD, CVM, and HH1) or Jersey sequenced bulls (JH1) had the core haplotype (not considering the disease allele itself). This ranged from zero, for JH1, to 159, for CVM. For all diseases except JH1, the core haplotype also occurred in other breeds (where these diseases have never been observed), though at very low frequency, and in only a small number, except for CVM.

Table 2. Length of core haplotype shared by all whole genome sequenced carriers of the disease (rare) allele for four lethal recessive diseases, number of non-carriers in which core haplotype is found, and number of other breeds in which core haplotype is found.

	BLAD	CVM	HH1	JH1
Number of carriers with whole genome sequence	6	30	7	12
Variants in core haplotype (shared by carriers)	302	93	437	633
Length of core haplotype (bp)*	40,362	21,020	57,173	48,608
Number of non-carriers in which core haplotype is found	4	159	1	0
Number of other breeds in which core haplotype is found	1	24	2*	0

*One of these was Danish red, which has Holstein introgressions

Given these results, we can start to speculate why the imputation of CVM genotypes is so poor, while for JH1, HH1 and BLAD imputation is more precise. The background haplotype in which the CVM mutation occurs, appears to be very common, even across breeds. It is likely that the CVM mutation occurred recently into this common haplotype background, such that there are otherwise identical haplotypes at reasonable frequency, without the mutation. This makes imputation, which is based on haplotype information, very challenging. In contrast, the JH1 mutation is imbedded in a longer haplotype which was likely at a lower frequency at the time the mutation occurred, such that carriers of the haplotype are also very likely to be carriers of the mutation as well. Parameters such as the frequency of the core haplotype into which the rare mutation occurred likely explain results from other studies as well, such as those of Bouwman *et al*. (2014), where reference sets for imputation which included multiple breeds improved accuracy of imputing a proportion of rare variants, but not others, compared to single breed reference sets.

Is there any way to improve the precision of imputing rare variants in light of the above? One of the first tasks is to reduce the error rate of genotyping variants from the whole genome sequence data – this complicates the identification of the core haplotype shared by carriers of the rare allele, and importantly might reduce the length of the core haplotype that can be confidently identified, which will reduce the accuracy of imputation (longer shared haplotypes between individuals lead to more precise imputation, eg Sargolzaei *et al*. 2014). Phasing errors are also important (phasing is necessary for imputation both in the sequenced animals and in the animals genotyped with 630K, and there could be errors in either), and are compounded by genotyping errors. So reducing genotyping errors could also improve the accuracy of phasing the data, which is desirable as any switch errors (false positive recombinations), if these are in the reference animals, will also reduce precision of imputation in the target animals. A practical way to remove some genotyping errors would be to run imputation for very rare variants within a breed, or combine LD information across closely related breeds (based on Fst for example)), only considering variants that segregate within the breed or group of breeds. This would reduce the number of variants (per breed), and therefore the opportunities for genotyping error, by 50% (Daetwyler *et al*. 2014). Information could then be accumulated across breeds.

ACKNOWLEDGEMENTS

The authors thank all members of the 1000 Bull Genomes Consortium for provision of data.

REFERENCES

Chadeau-Hyam M., Hoggart C., O'Reilly P., Whittaker J., De Iorio M. and Balding D. (2008) *BMC Bioinformatics* **9**: 364.

Erbe M., Hayes B.J., Matukumalli L.K., Goswami S., Bowman P.J., Reich C.M., Mason B.A. and Goddard M.E. (2012) *J. Dairy Sci.* **95**: 4114.

Shuster D.E., Kehrli M.E. Jr, Ackermann M.R., Gilbert R.O. (1992) *Proc Natl Acad Sci U S A*. **89**:9225.

Thomsen B., Horn P., Panitz F., Bendixen E., Petersen A.H., Holm L.E., Nielsen V.H., Agerholm J.S., Arnbjerg J., Bendixen C. (2006) . *Genome Res*. **16**:97.

Adams H.A., Sonstegard T., VanRaden P.M., Null D.J., Van Tassell C., *et al*.. (2012) *Plant and Animal Genome Meeting*, Poster P0555, June 14–18 2012, San Diego.

Sonstegard T.S., Cole J.B., VanRaden P.M., Van Tassell C.P., Null D.J., *et al*. (2013) *PLoS ONE* **8**: e54872.

Sargolzaei M., Chesnais J.P., Schenkel F.S. (2014) *BMC Genomics*. **15**:478.

Howie B., Fuchsberger C., Stephens M., Marchini J., Abecasis G.R. (2012) *Nat Genet* **44**:955

Daetwyler H.D., Capitan A., Pausch H., *et al*. (2014) *Nat Genet*. **46**:858.

Bouwman A.C., Veerkamp R.F. (2014) *BMC Genet.* **15**:105.

1000 Genomes Project Consortium, Abecasis G.R., Auton A., *et al*. (2012) *Nature.* **491**:56.

Proc. Assoc. Advmt. Breed. Genet. 21: 45-48

FROM SHEEP SNP CHIPS, GENOME SEQUENCES AND TRANSCRIPTOMES VIA MECHANISMS TO IMPROVED SHEEP BREEDING AND MANAGEMENT

B.P. Dalrymple[1], V.H.Oddy[2], J.C. McEwan[3], J.W. Kijas[1], R. Xiang[1], J. Bond[2], N. Cockett[4], K. Worley[5], T. Smith[6] and P.E. Vercoe[7]

[1] CSIRO Agriculture Flagship, Queensland, Australia
[2] NSW Department of Primary Industries, Beef Industry Centre, New South Wales, Australia.
[3]AgResearch, Otago, New Zealand
[4] Utah State University, Utah, USA
[5] Human Genome Sequencing Center, Baylor College of Medicine, Texas, USA
[6] UDSA MARC, Nebraska, USA
[7] Institute of Agriculture, University of Western Australia, Western Australia, Australia

SUMMARY

SNP chips are transforming animal breeding; low cost "assay-by-sequencing" methodologies and high quality reference genome sequences provide the opportunity for further significant improvement in both breeding and management. The Functional Annotation of ANimal Genomes (FAANG) consortium is applying methods developed by the human ENCODE project to annotate the genomes of livestock (sheep, cattle, pigs, etc.) with functional information including the probability that variation at a particular nucleotide has a causal role in any phenotype. We will contribute the detailed annotation of the transcriptome of the gastrointestinal tract of sheep to FAANG. We will undertake an integrated analysis of the variation in: genome sequence, transcription, gastrointestinal tract phenotypes and the environment across ~100 animals. This will be combined with analysis of a developmental time course of the gastrointestinal tract transcriptome from 30 days post conception to weaning, and an in-depth analysis of the gastrointestinal tract transcriptome from the new reference sheep, a North American Rambouillet. From this, and public FAANG data, we will estimate the probability that variation in a particular nucleotide has an impact on gastrointestinal phenotypes of interest (methane, nutrition, infection, microbial population) and identify the biological processes underlying the phenotype. This information will inform breeding schemes, identify management options and define phenotypes more precisely.

INTRODUCTION

Over the last ten years there has been a paradigm shift in the use of genetic markers in animal breeding. Microsatellites have been very quickly replaced by Single Nucleotide Polymorphisms (SNPs). For sheep the first whole genome SNP data was generated using a 1536 SNP platform (Kijas *et al.* 2009). This was soon followed by the Ovine SNP50 BeadChip with 50K SNPs (Kijas *et al.* 2012) and more recently a high density SNP-Chip with more than 600K SNPs (Kijas *et al.* 2014). In addition, targeted small SNP chips have been designed for specific purposes, such as parentage testing and use in industry breeding programs (Heaton *et al.* 2014). SNP genotyping information from large numbers of individuals is being applied in sheep breeding programs (Auvray *et al.* 2014, Moghaddar *et al.* 2014). However, the vast majority of the SNPs are still markers in linkage disequilibrium with the causative variation, not the causative variation themselves. Accurate identification of the causative variants would increase the accuracy of the prediction equations, by removing the linkage uncertainty. In addition, SNPs are not the only variations in the genome with a causal role in phenotype variation, for example duplications and rearrangements are involved in agouti (Norris and Whan 2008) and weight of lamb weaned

(Gonzalez *et al.* 2013). Copy number variations (CNVs) of a range of sizes are common in the sheep genome (Jiang *et al.* 2014). Some, but not all, of this variation is captured by linked SNPs (Gonzalez *et al.* 2013).

The search for causative mutations has identified a small number of large effect in sheep, including myostatin (Clop *et al.* 2006) and Callipyge (Smit *et al.* 2003). In both cases the causative mutations are not in the coding region of genes, but are in associated regulatory sequences, a new micro RNA-binding site (myostatin) and a methylated control region (Callipyge). Whilst the effect of variations in coding sequences on the function of the gene products can be predicted fairly reliably, this is not the case for variations in non-coding sequences such as long non-coding RNAs (lncRNAs) and regulatory sequences to which transcription factors bind. Experimental validation of causative mutations can only be justified for mutations of large effect. For most phenotypes many genes of small effect are involved and high throughput data generation followed by computational analysis is the only realistic way to approach the genome-wide identification of causal mutations. The first major barrier to effective prediction is that the role of the majority of individual nucleotides in the genome of production animals is not known. One of the major goals of the human and model organism ENCODE projects is to identify the role, or not, of each nucleotide in the genome (Dunham *et al.* 2012). To do this these projects have focussed on a small number of approaches, generally “assay-by-sequencing” methodologies including: in-depth transcriptomics, methylation, chromatin accessibility and conformation, transcription factor binding sites etc. Thus, across the vast majority of the genome each nucleotide can be annotated for a number of attributes: in a transcription factor binding site, transcribed, in an exon, in a splice site, in open or closed chromatin, etc. For each of these attributes the effect of variation on the role of the nucleotide in the functional element can be estimated (Gulko *et al.* 2015). This prediction is phenotype independent. Subsequently in studies of the association between variation in the genome and variation in the phenotype and the calculation of predictive equations, the probability that variation in a particular nucleotide will affect a downstream process can be included into the equations, and genetic relationship matrices (GRM) can be built using causative sites. The utility of this approach has been demonstrated in preliminary analyses (Gusev *et al.* 2014, Koufariotis *et al.* 2014). The FAANG consortium has been established to coordinate the international projects for the annotation of the roles of the nucleotides within the genomes of the major production animal species using the methodologies validated in the ENCODE projects (Andersson *et al.* 2015) (Figure 1).

However, association studies using the FAANG generated datasets will also inform our understanding of biological processes underlying a phenotype, by providing an estimate of the probability of a particular variation in the genome sequence affecting the phenotype of interest. The increased understanding of the biological processes will also be used to improve the management of the animals to reach their genetic potential. In addition, understanding the biological processes underlying the phenotype may enable us to define phenotypes better, reducing complex phenotypes to a series of simple phenotypes based on different biological processes (Figure 1).

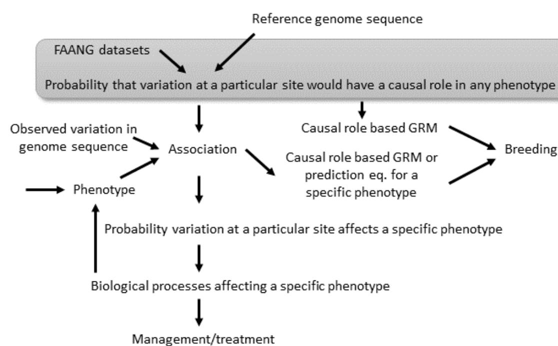


Figure 1. Pipeline for FAANG annotation (highlighted) and delivery.

The gastrointestinal tract (GIT) is the

major source of nutrients in animals and an important source of their waste, such as methane (Johnson and Johnson 1995); appropriate function of the GIT is essential for the efficient production of animals and their products. A more detailed understanding of the genes and gene products (and their regulation) contributing to the development and function of the GIT, and hence the biological processes involved, will facilitate the development of new breeding strategies, methodologies for feeding animals, and managing the function of the GIT. Successful implementation of these will increase production efficiency and reduce waste products/kg of meat, milk, wool etc.

MATERIALS AND METHODS

In preparation for FAANG a new version of the sheep reference genome sequence is being assembled (Oar v4) using the long read PacBio sequencing technology. In addition, we will initially build scaffolds *de novo* using Hi-C (Burton *et al.* 2013), followed by using the ovine BAC library (Dalrymple *et al.* 2007), and the SNP-based sheep linkage and RH maps (Jiang *et al.* 2014). The individual being sequenced, a North American Rambouillet, will also be the source of reference tissue samples for the FAANG consortium assays. The Oar v3.1 assembly (Jiang *et al.* 2014), based on a male and a female Texel, will be gap-filled and errors corrected using a low coverage of PacBio sequencing of the male Texel to generate Oar v3.2.

Samples along the GIT, salivary gland, reticulum, sacs of the rumen, omasum, abomasum, duodenum, cecum, colon and rectum, have been collected from 63 Australian sheep (ewes) with a diversity of origins (Merino and Suffolk, Border Leicester, Dorset cross bred animals) and from 48 NZ sheep (ewes and rams) from a high/low methane selection line, also with a diversity of origins (Pinares-Patino *et al.* 2013). Samples from a time course of the development of the whole GIT of Merino sheep from 30 days post conception to weaning have been collected. Sampling times were selected based on the development of the sheep GIT (Franco *et al.* 1992). We will develop a detailed description of the transcriptome of the GIT of sheep by undertaking Illumina RNA-Seq and PacBio Iso-Seq on mRNA, lncRNA and small RNA (miRNA, snoRNA etc.) of the sheep GIT tissues. Gene and transcript models of protein and non-coding RNAs will be built using both assembly guided and *de novo* methodologies. The significantly expanded GIT-relevant transcriptome will identify transcript isoforms currently poorly represented in the sheep transcript models. The focus of the targeted manual annotation of transcripts will be gene products identified as likely to play key roles in the development and function of the GIT. We will use the correlation of the expression of transcripts with each other and with development and productivity GIT phenotypes measured (and microbial samples from the GIT collected) as part of the Australian Department of Agriculture funded "Host control of methane emissions from sheep" project to create gene/transcript networks, and to identify sets of genes and their transcripts informative of key biological processes in the sheep GIT. We will also map the changes in the processes during development of the GIT. We will use the sets of transcripts to identify transcription factors and other regulatory molecules likely to be involved in the regulation of key processes in the sheep GIT. All the data will be generated following the FAANG standard operating protocols and will be contributed to the FAANG consortium. We will promote and facilitate the use of the data by making it publicly available prior to publication in accordance with standard data sharing protocols and by providing online support for users.

RESULTS AND DISCUSSION

Analysis of the sheep genome and transcriptome has identified a strong relationship between the rumen and cornified and keratinized tissues such as the skin and ruminant specific genes encoding proteins predicted to be involved in the cornification of the rumen epithelium (Jiang *et al.* 2014). A first generation sheep GIT transcriptome atlas is currently being constructed using

data generated as part of the sheep reference genome project. Initial analysis of the rumen gene expression from 24 NZ ewes has been undertaken demonstrating that the expression of genes relating to the cornified epithelium is dynamic, probably due to dietary effects (Ruidong et al., in preparation).The initial output of this work will be an in-depth and detailed catalogue of the transcriptome of the sheep GIT tissues increasing the number of alternate splice variants from an average of one per gene to more than four. Annotation accuracy and coverage of transcript isoforms is expected to be high for mRNAs and slightly lower for short RNAs, such as miRNAs. Conversely the discovery rate of new lncRNAs is expected to be high and of new protein coding genes is expected to be low.The developmental gene expression atlas of the GIT of sheep and the integration of other FAANG datasets with the atlas will support global research activities into the development and function of the GIT and how putative causative SNPs and other genomic variation affects the efficiency of animal growth, methane production, parasite resistance and other traits critical to the profitability and sustainability of livestock production. For example it is expected that reticulo-rumen, and possibly salivary gland gene expression will be associated with differences in rumen morphology and digesta flow and in turn methane production, and that abomasal and duodenal gene expression patterns will inform how sheep differ in parasite resistance.

CONCLUSIONS

From the research described above causative SNPs and CNVs weighted for phenotypic impact on GIT function will be available for inclusion in GRMs and prediction equations in breeding schemes. It is likely that these analyses will also identify biological pathway based phenotypes which may enable new approaches to the selection of animals for production traits to be identified. In contrast the pathway to the utilization of the outputs for improved management of sheep is much more poorly defined and likely to be more challenging.

REFERENCES

Andersson L., Archibald A., *et al.* (2015). *Genome Biol.* **16**: 57.
Auvray B., McEwan J.C., *et al.* (2014). *J. Anim. Sci.* **92**: 4375.
Burton J.N., Adey A., *et al.* (2013). *Nat. Biotechnol.* **31**: 1119.
Clop A., Marcq F., *et al.* (2006). *Nature Genet.* **38**: 813.
Dalrymple B.P., Kirkness E.F., *et al.* (2007). *Genome Biol.* **8**: 20.
Dunham I., Kundaje A., *et al.* (2012). *Nature* **489**: 57.
Franco A., Regodon S., Robina A. and Redondo E. (1992). *Am. J. Vet. Res.* **53**: 1209.
Gonzalez M.V., Mousel M.R., *et al.* (2013). *PLoS One* **8**: 9.
Gulko B., Hubisz M.J., Gronau I. and Siepel A. (2015). *Nature Genet.* **47**: 276.
Gusev A., Lee S.H., *et al.* (2014). *Am. J. Hum. Genet.* **95**: 535.
Heaton M.P., Leymaster K.A., *et al.* (2014). *PLoS One* **9**: 10.
Jiang Y., Xie M., *et al.* (2014). *Science* **344**: 1168.
Johnson K.A. and Johnson D.E. (1995). *J. Anim. Sci.* **73**: 2483.
Kijas J.W., Townley D., *et al.* (2009). *PLoS One* **4**: 13.
Kijas J.W., Lenstra J.A., *et al.* (2012). *PLoS. Biol.* **10**: 14.
Kijas J.W., Porto-Neto L., *et al.* (2014). *Anim. Genet.* **45**: 754.
Koufariotis L., Chen Y.P.P., Bolormaa S. and Hayes B.J. (2014). *BMC Genomics* **15**: 16.
Moghaddar N., Swan A.A. and van der Werf J.H.J. (2014). *Anim. Prod. Sci.* **54**: 544.
Norris B.J. and Whan V.A. (2008). *Genome Res.* **18**: 1282.
Pinares-Patino C.S., Hickey S.M., *et al.* (2013). *Animal* **7**: 316.
Smit M., Segers K., *et al.* (2003). *Genetics* **163**: 453.

Proc. Assoc. Advmt. Breed. Genet. 21: 49-52

PLEIOTROPIC MULTI-TRAIT GENOME-WIDE ASSOCIATION REVEALS PUTATIVE CANDIDATE GENES FOR FATTY ACID COMPOSITION IN AUSTRALIAN SHEEP

S. Bolormaa[1,2], J.H.J. van der Werf [2,3], B.J. Hayes[1,2,4], M.E. Goddard[1,5], and H.D. Daetwyler[1,2,4]

[1]Department of Economic Development, Jobs, Transport and Resources, Bundoora, Australia
[2]CRC for Sheep Industry Innovation, Armidale, Australia
[3]University of New England, Armidale, Australia
[4]La Trobe University, Bundoora, Australia
[5]University of Melbourne, Parkville, Australia

SUMMARY

The aims of this paper were to detect SNP and genes that affect the fatty acid composition and other carcass and meat quality traits in sheep. We performed genome-wide association studies (GWAS) for 56 traits including carcass weights, muscling, fatness, tenderness, meat color, mineral content, and fatty acid composition on 10,613 animals genotyped for 510,174 SNPs. The use of a meta-analysis to combine information from the 56 traits increased the power to detect QTL compared with the single trait analyses. We found pleiotropic QTL, which appear to cluster into 5 functional groups based on their trait effects. Candidate genes were identified in the groups that have functions consistent with the biology of the traits.

INTRODUCTION

The fatty acid (FA) composition (FAC) of meat products has received considerable attention for its significance to human health. The FAC of meat is influenced by various environmental effects such as diet (e.g. Suzuki *et al*. 2007), the level of fatness, and genetic factors. Changing FAC through selection could decrease the saturated FA (SFA) content of meat and perhaps improve human health. Breed or genotype differences in the FAC have been reported, even after correction for fat level (De Smet *et al*. 2004). FAC can be described by a combination of traits, but GWAS are usually performed one trait at a time, which may reduce the power to detect mutations that affect the trait complex. In fact, mutations that affect FAC may also affect other quality and carcass traits, therefore a multi-trait GWAS analysis is expected to be more powerful to assess the effect of a mutation on multiple related traits. Multi-trait analysis of linkage experiments has been reported to increase the power to detect QTL (e.g. Knot and Haley. 2000). Bolormaa *et al*. (2014) found increased power to detect QTL by a meta-analysis that combined the results from GWAS for 32 individual traits in beef cattle. The objectives of this study were to identify genes that affect FAC and other meat quality and carcass traits using a multi-trait GWAS in Australian sheep.

MATERIALS AND METHODS

SNP and phenotype data. In total, 510,174 single nucleotide polymorphisms (SNP) were genotyped or imputed from lower density SNPs. The SNP genotype and phenotype data was collected by the CRC for Sheep Industry Innovation (Sheep CRC) and SheepGENOMICS projects. The SNP were obtained from two different SNP arrays: the Illumina 600k (HD) and 50k Ovine SNP chips (Illumina Inc., San Diego, CA, USA). All SNP were mapped to the OAR 3.1 build of the ovine genome sequence assembled by the SNPchiMp v.3 (Nicolazzi *et al*. 2015). Details on genotyping, editing and imputation of the Sheep CRC 50k data set has been described by Daetwyler *et al*. (2012). The imputation from 50k to HD was done using Fimpute (Sargolzaei *et al*. 2014). Out of 22,684 animals with 50k and 1,735 HD genotypes, the accuracy of imputation between 50k and HD genotypes for non-50k SNPs was 0.9871. A total of 10,613 sheep (from multiple breeds and crosses) HD genotyped were measured for up to 56 meat related traits (carcass

weight, fatness, muscling, tenderness, meat colour, pH level, and FAC). A complete description of the design, methods and analyses of carcass and meat quality assessments is given by Mortimer *et al*. (2014).

Statistical analysis. A GWAS was carried out for each trait using ASReml software (Gilmour. 2009) to fit a mixed model: trait ~ mean + fixed effects + SNP_i + animal + dam + sire by flock interaction + error; with animal, dam (permanent environment), and sire by flock interaction, and error fitted as random effects including relationships between animals. All models included dataset, management group, flock, date of observation, drop year, sex, birth type, and rear type as fixed effects. The FA traits were corrected for intramuscular fat content. The individual trait results were combined using the meta-analysis described by Bolormaa *et al*. (2014). To avoid identifying a large number of closely linked SNPs, whose association with traits is due to the same QTL, only the most significant SNP ($P<10^{-5}$) from each 1Mbp interval was retained for validation. In a further restriction, a maximum of two or three SNPs on the same chromosome were selected, and only if they clearly represented different QTL based on the multi-trait test. In this way, 23 "Lead SNP" representing different QTL across the genome were selected. For each lead SNP, a linear index of the 56 traits, that was maximally correlated with the Lead SNP genotypes in the reference population (4/5 of total animals) was constructed and used to validate SNP effect in the other 1/5 of the data, as described by Bolormaa *et al*. (2014). Clustering analysis based on a correlation matrix among the 23 Lead SNP genotypes was used to assign the lead SNPs into 5 groups such that the SNPs within a group had a similar pattern of effects across the 56 traits. To find additional SNPs with a similar pattern of effects to each of the 23 lead SNPs, the linear index for the corresponding Lead SNP was used as a new "phenotype" in GWAS (which fitted the 23 Lead SNP themselves as fixed effects). Genes that occurred within 30 kb of the significant SNPs were annotated using UCSC Genome Bioinformatics (genome.ucsc.edu) and Ensembl (www.ensembl.org/biomart/). The statistical enrichment of the biological function or pathways in groups of genes were checked using the STRING program (Franceschini *et al*. 2013).

RESULTS AND DISCUSSION

Using the multi-trait analysis, 586 SNPs were significant ($P < 5\times10^{-7}$), corresponding to a false discovery rate of 0.04%, and this was better than for any individual trait. When traits were analyzed individually, for only 10 out 56 traits the FDR was less than 2.5%. Many highly significant SNPs from the multi-trait analyses were found within narrow regions on Ovine autosomal chromosomes (OAR) 2, 3, 5, 6, 11, 12, 14, 18, 20, and 26. Many of the significant SNPs in both single trait and multi-trait analyses were closely linked and could be associated with the same QTL. When only the most significant SNPs in each Mb interval were retained from the multi-trait analysis of the discovery dataset for validation purpose, 98 SNPs were significant at $P < 10^{-5}$. In the validation population, all 98 SNPs had an effect in the same direction as in the discovery population and 35 were significant ($P < 0.05$). One of the best single-traits was DRESSING% where 9 of 31 significant ($P < 10^{-5}$) SNPs found in reference population were significant at $P < 0.05$ in the validation population. Therefore the multi-trait analysis detected associations with higher reliability than single-trait analysis.

The multi-trait analysis was particularly successful in detecting pleiotropic QTL. Genes that operate in the same pathway might be expected to show the same pattern of pleiotropic effects. The cluster analysis of the 23 Lead SNP revealed 5 loosely-defined groups of SNPs where the SNPs within a group had a similar pattern of effects across traits: 1) changing skeletal or carcass size; 2) changing fat deposition (SFA profile and fatness); 3) influencing meat eating quality; 4) affecting meat colour; 5) a poorly defined group affecting only specific trait(s) (e.g. changing only glycogen level, omega-3 or omega 6 FA). The GWAS using Lead SNP linear indices identified 687 significant SNPs ($P < 5\times10^{-7}$) assigned to each of the 5 groups as follows: 1) 518; 2) 30; 3) 99,

4) 25; and 5) 15.

Figure 1. Correlation matrix between the effects of lead SNPs and their linear index SNPs (chromosome_position_annotated gene name) within Group 2.

Group 2 initially consisted of four strongly correlated Lead SNP on OAR 6 (OAR6_15.2Mb), 11 (OAR11_13.3Mb and OAR11_49.9Mb), and 26 (OAR26_13.9Mb). These 4 SNPs have alleles that increase the concentration of SFAs with carbon chain of C16, C14, C12, and C10 (palmitic, myristic, lauric, and capric acids, respectively) and may decrease stearic acid profile (C18:0) (Table 1). There was also a tendency for SNP alleles that increased saturated FAC to increase fatness. If each QTL group represented a particular physiological pathway, one might expect the genes that map near the QTL of a group to show some similarity of function. The 4 Lead SNPs in Group 2 were expanded with 7 additional SNP from the linear index GWAS and the clustering was repeated within group (Figure 1). These additional 7 SNPs were chosen to represent different QTL or as multiple candidates for the same QTL where more than one gene was a suitable candidate. The 11 SNPs were clustered into two subgroups (Figure 1) and they map near to or within genes affecting FA traits. Sub-group 1 consists of SNPs whose effects were moderately to highly correlated (r > 0.4) and tagged genes such as *FASN*, *MLXIPL*, *EVOLV6*, *ACACA*, and *SYNRG*. The other more loosely correlated sub-group included the *ACSL1*, *ISYNA1*, *SGK2*, and *AGPAT9* genes (Figure 1). The former group includes genes which play a role in the synthesis of fatty acid formation and the latter includes genes with a role in (glycero)lipid biosynthesis. The *MLXIPL* protein activates carbohydrate response element motifs in the promoters of triglyceride synthesis genes. Studies in gene expression (e.g. De Jager *et al.* 2013) have reported that *ACACA*, *FASN,* and *DGAT2* may directly influence IMF percentage in sheep and cattle. Dervishi *et al.* (2011) noted that changes in FA profile due to feeding systems implicate changes in the mRNA expression level of genes related with fat metabolism. Thus, group 2 does represent a set of genes with known function in fatty acid and lipid synthesis. This was confirmed by KEGG and GO enrichment analysis. There were some enriched interactions at medium confidence level (P = 6.9×10^{-6}) between some Group 2 genes. GO and KEGG analysis found some evidence for

statistically significantly enriched in the biological function or pathways in our FA genes. For example, according to KEGG and GO terms, two proteins (FASN and ACACA) were involved together in FA biosynthesis (Bonferroni $P = 3.8\times10^{-4}$), 5 proteins (FASN, ACACA, ACSL1, AGPAT9, and ISYNA1) in metabolic pathways (Bonferroni $P = 7.6\times10^{-3}$), and 4 proteins (FASN, ELOVL6, ACSL1, and ISYNA1) in lipid biosynthesis process (Bonferroni $P = 3.8\times10^{-2}$).

Although the SNPs within a group share some features they may also be associated with seemingly quite different traits. Two SNPs from Group 3 (near to genes *MRPS25* and *CYP27A1*), one SNP from Group 4 (near *GLTPD1*) and two from Group 5 (within genes *PNPLA3* and *FADS2*) affected the concentration of polyunsaturated FA (Table 1).

Table 1. Effect of a subset of significant multi-trait SNPs in the individual fatty acid traits (signed t-values >1 are shown)

			Fatty acid with chain of										
#[1]	OAR[2]	POS[3]	C22: 6n-3	C22: 5n-3	C20: 5n-3	C20: 4n-6	C20: 3n-6	C18: 2n-6	C18: 0	C16: 0	C14: 0	C12: 0	C10: 0
2	26	13.9	2.8			1.5	-1.2	2.7				-4.5	-3.3
2	11	13.2	1.0		1.1	1.3	2.6		-2.5	3.6	4.9	2.2	7.2
2	11	49.9							-1.6	4.9	8.4	3.3	3.8
2	6	15.2	-1.8			-1.4	-1.2		-1.8	6.0	3.7	3.0	3.0
3	19	57.1	2.4	2.8	2.4	1.9	2.2		-1.0				-2.1
3	2	219	3.6	1.0	4.1	5.4	2.8	3.7		-1.6	-2.9	-3.3	-3.0
4	12	49.6	-6.8	-5.8	-8.2	-4.6	-5.7	-6.1	1.4				
5	3	21.8		3.5	3.3	-1.6	2.4			-2.0		2.9	1.2
5	21	39.7	1.5			3.1	8.2	-1.8					

[1]Group; [2]Chromosome; [3]Ovine chromosome position in megabases.

The pattern of effects of each QTL studied here indicated that some may be more useful for selection than others depending on the breeding goal. Some QTL have an allele with desirable effects on more than one trait and appear to be good targets for selection. For instance, the QTL on OAR 2 had allele that increases tenderness, improves meat colour, increases myoglobin, glycogen, and omega-FAs and decreases long chain saturated FAs, which is a highly valuable pattern. Selection for this allele would be beneficial in sheep intended for most markets because lamb prices reflect colour, tenderness, palatability, and juiciness.

ACKNOWLEDGEMENTS

We thank Klint Gore and Ken Geenty for managing the CRC information nucleus database and the many staff involved at the CRC and SG sites across Australia.

REFERENCES

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., et al. (2014) *PLoS Genet.* **10**:e1004198
Daetwyler H.D., Swan A., van der Werf J.H.J., Hayes, B.J. (2012) *Gen. Sel. Evol.* **44**:33.
De Jager N., Hudson N.J., Reverter A., Barnard R., *et al*. (2013) *J. Anim Sci.* **91:**1112–1128.
De Smet S., Raes K., Demeyer D. (2004) *Anim. Res.* **53**:81.
Franceschini A., Szklarczyk D., Frankild S., *et al*. (2013) *Nucl. Acids Res.* **41**:D808-D815.
Knott S. Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. (2009) HP1 1ES, UK.
Knot A. and Haley C. S. (2000) *Genetics.* **156**:899.
Mortimera S.I., van der Werf J.H.J., Jacobd R.H., *et al*. (2014) *Meat Sci.* **96**:1016.
Nicolazzi N.L., Caprera A., Nazzicari N., Cozzi P., *et al.* (2015) *BMC Genomics* **16:**283.
Sargolzaei M., Chesnais J., Schenkel F. (2014) *BMC Genomics*. **15**:478.
Suzuki S., Ishikawa S., Arihara K., Itoh M. (2007) *J. Anim. Sci.* **78**:293.

Proc. Assoc. Advmt. Breed. Genet. 21: 53-56

FINE-MAPPING SINGLE NUCLEOTIDE POLYMORPHISMS ON *BOS TAURUS* CHROMOSOME 26 AFFECTING ADIPOSE MYRISTIC ACID

D. Lu[1], S.M. Hickey[2], N.G. Cullen[2] and S. Miller[1] Invermay
[1] Agricultural Centre, Mosgiel 9053, New Zealand
[2]Ruakura Research Centre, Hamilton 3240, New Zealand

SUMMARY

It may be desirable to change the genetic selection program of cattle to make them produce less myristic fatty acid (alternatively C14:0) in milk and adipose fat, providing food products to consumers with enhanced health attributes. As phenotypic measures on an industry-wide basis are not practical, a genomic test related to levels of C14:0 would provide a viable method to enhance animal selection programs. Therefore it is important to identify genome segments and nucleotides that control the trait. The study reported herein was able to locate a region of ~250Kb on chromosome 26 that accounted for 26.47% of the phenotypic variation in C14:0. Two of eight haplotypes of this region were found to reduce C14:0 significantly in subcutaneous fat of beef cattle.

INTRODUCTION

Dietary fatty acids, especially C14:0, have a major influence on lipoprotein concentrations in human plasma (Katan *et al*. 1994; Adamsson *et al*. 2014), which in turn affects cardio-vascular health (Mozaffarian *et al*. 2005). Those fatty acids that increase undesirable cholesterol in humans, are mainly derived from milk fat (Gunstone *et al.* 1994), and beef meat (Youssef *et al.* 2012). Therefore reduction of those fatty acids in milk and beef is desirable.

From a genetics perspective, the proportion of C14:0 to total lipid amount shows evidence of being under genetic control (Tait *et al.* 2008; Bouwman *et al.* 2011). Its heritability in beef adipose fat was reported to be 0.50 (Tait *et al.* 2008). This means it is possible to select cattle for the reduction of C14:0 in meat. Selecting animals via traditional progeny testing is time consuming due to a long generation interval and the complexity of measuring C14:0, would result in slow genetic gain. Current DNA technology can help facilitate the genetic improvement process *via* marker assisted selection or genomic selection. For this to happen, it is important to identify single nucleotide polymorphisms (SNP) that are associated with levels of C14:0.

Morris *et al.* (2007; 2010) identified quantitative trait loci (QTL) on *bos taurus* autosomes (bta) 15, 19, 26, 27 and 29 for C14:0 in adipose fat. However, the QTL regions covered long segments of chromosomes, for example, 18-29cM on bta 26 (Morris *et al.* 2010). Using higher density genetic markers would allow for fine mapping genomic regions in association with the trait.

The study reported here was aimed at refining the QTL regions reported by Morris *et al.* (2007; 2010) for C14:0 in a New Zealand experimental population of Jersey-Limousin backcrosses.

MATERIALS AND METHODS

Animals and Phenotype. Records of fatty acid profiles were available on 406 backcrossed Jersey (J) x Limousin (L), which were sired by three JxL bulls *via* artificial insemination with J and L dams. The animals were raised on pasture and slaughtered at 22-28 months of age. Subcutaneous fat from over the *longissimus dorsi* muscle was used to extract fatty acids and nine (C14:0, C14:1, C15:0, C16:0, C16:1, C17:0, C18:0, C18:1 and C18:2) were measured, and presented as percentage of the total of the nine fatty acids. The mean for C14:0 was 3.41±0.48. More details of the animals and phenotype measurement are described by Morris *et al.* (2010).

Genotypes. Genotypes were available on only 160 heifers and 106 steers born in 1996-1997. These animals formed three half-sib families with 74, 94 and 98 progeny. The genotyping was

performed on DNA extracted from blood, ear tissues, and meat samples. A total number of 54,609 SNP were typed across the bovine genome by Delta Genomics Laboratory (Edmonton, Alberta, Canada), using the Illumina BovineSNP50 Beadchip (Illumina Inc., San Diego, USA). Animals with a call rate less than 90% were removed. Quality control filtered out SNP that had a minor allele frequency less than 1%, or a spurious location, or a GC score less than 0.15. Missing genotypes were imputed using FImpute v2.2 (Sargolzaei *et al.* 2014), which makes use of both family and population information. A subset of 39,988 SNP on 29 autosomes was used for subsequent analyses.

Linkage Disequilibrium. Pairwise linkage disequlibrium (LD) was estimated using $r^2 = \frac{D^2}{f(A).f(a).f(B).f(b)}$ (Hill and Roberson 1968), with $D = f(AB) - f(A).f(B)$, where $f(AB)$ is the estimated frequency of haplotype AB using the observed genotype frequency (McVean 2007) and assuming Hardy-Weinberg equilibrium, $f(A), f(B), f(a), f(b)$ being the observed frequencies of alleles A, B, a, b, respectively. The metric r^2 was computed using software Snppld v1.0 (Sargolzaei 2010).

Statistical analyses. Genome-wide association analysis was carried out to identify chromosome segments that potentially harbour SNP or haplotypes in strong association with C14:0. The trait was adjusted for fixed effects, including breed of dam (J or L), farm of birth (n=3), birth type (single or twin) within breed of dam, slaughter group (sex and year included), sire family, then fitted in the following model $y_{ij} = \mu + \beta X_{ij} + e_{ij}$, where μ is the overall mean, y_i the adjusted C14:0 for animal i, β the regression coefficient for genotype X $\{0,1,2\}$ at locus j, e_{ij} the residual.

Haplotype phase. Haplotypes were reconstructed for only chromosome regions that had high LD between pairs of SNP and contained SNP highly associated with C14:0, using Beagle software v3.3.2 (Browning and Browning 2007). Haplotype effects were estimated from the statistical model mentioned above by replacing genotypes with haplotypes.

RESULTS AND DISCUSSION

Genome-wide association. Significance levels of SNP from the association analysis are presented in Figure 1. The two peaks were observed at *rs41921177* (19:51326750; FDR= 1.03E-6) and *rs110857021* (26:21832456; FDR = 1.03E-6). Within 500kb of *rs41921177* are four other SNP significant at FDR<5% and gene *FASN* (fatty acid synthase), which is reported to be associated with fatty acid composition in beef muscle (Zhang *et al.* 2008; Yokota *et al.* 2012), and milk fat (Roy *et al.* 2006). Two of the four other significant SNPs were within 50kb of gene *FASN*. Within 500kb of *rs110857021* are 14 other SNP significant at FDR<1% and gene *Stearoyl-CoA Desaturase,* which is reported to affect fatty acid composition of adipose tissue in beef cattle (Brooks *et al.* 2011; Yokota *et al.* 2012; Costa *et al.* 2013), and of milk fat (Rincon *et al.* 2012).

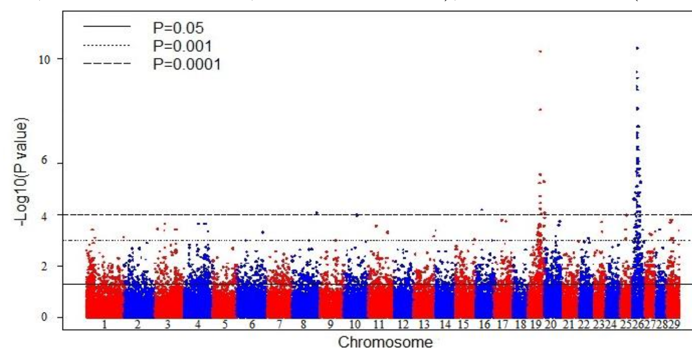


Figure 1. Significance levels of SNP from the association with myristic fatty acid.

Linkage disequilibrium was estimated for pairs of all SNP on bta 19 and 26. Figure 2 displays the LD heatmap for chromosome segments of approximately 500kb on each side of *rs41921177* and *rs110857021*. Single nucleotide polymorphism *rs41921177* on bta 19 did not appear to have strong LD with its surrounding SNPs, while SNP *rs110857021* on bta 26 was highly correlated ($r^2 \geq 0.6$) with five other SNPs immediately adjacent to it. This finding supports the number of SNPs found significantly associated with C14:0 in these two chromosomal regions. Regions around *rs41921177* might need denser SNPs to capture higher LD, enabling higher confidence in identifying nucleotides that are causative for differences in C14:0. The region of six consecutive SNPs (~250kb) on bta 26 was further investigated, using phased haplotypes.

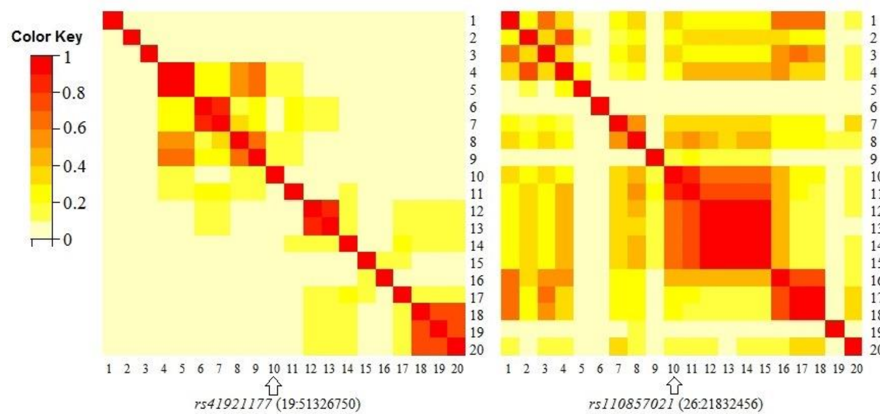


Figure 2. LD heatmap for segments of chromosomes 19 and 26.

Haplotype analysis. Eight haplotypes were found out of the six SNP region mentioned above. Their frequency is presented in Table 1. Four haplotypes with very low frequency (<1%) were grouped together and given code hap1. Five haplotypes (hap1-5) were used in a linear regression to estimate haplotype effect (Table 2).

Table 1. **Distribution of haplotypes on bta 26**

Haplotype (Code)	Sire1	Sire2	Sire3	Freq (%)
CACGGA (hap1)	1	0	4	0.94
CATAAC (hap2)	41	25	67	25.00
CCCAAC (hap1)	0	0	1	0.19
CCCGGA (hap3)	4	6	6	3.00
TACGGA (hap1)	0	0	1	0.19
TCCAGA (hap1)	0	0	1	0.19
TCCGGA (hap4)	96	151	105	66.17
TCTAAC (hap5)	6	6	11	4.32

Table 2. **Haplotype effect**

Hap	Effect	P value
hap4	0	N/A
hap1	-0.07 ± 0.27	> 0.05
hap2	-0.68 ± 0.09	< 0.001
hap3	-0.63 ± 0.22	< 0.005
hap5	-0.03 ± 0.19	> 0.05

In the analysis, effects of hap1, hap2, hap3, hap5 were contrasted against the effect of hap4. Hap4, as the most frequent haplotype among the studied animals, and a common haplotype among the three sires, appeared to be associated with the highest percentage of C14:0 in adipose fatty acids. Hap2 was the second most frequent haplotype, and appeared to be associated with a reduction in C14:0. Hap3 differed from hap4 only at locus *rs110857021*, where *T* in hap4 was replaced by *C* that subsequently caused a significant reduction in C14:0 ($P<0.005$). The ~250kb segment of bta 26 accounted for 26.47% of the phenotypic variation in C14:0.

CONCLUSION

Chromosome regions were refined, which were previously reported to be associated with C14:0. A very high LD segment of ~250kb on bta 26 was located accounting for 26.47% of the variation in C14:0 in the research animals. Two haplotypes (hap2 and hap3), which had a combined frequency of 28% and caused a reduction of myristic fatty acid, should be tested for association with other economically important traits, and tested in other cattle breeds or populations.

ACKNOWLEDGEMENTS

The NZ-half of the trial run in conjunction with the University of Adelaide was funded by New Zealand Foundation for Research, Science and Technology, and by J. S. Davies Bequest to the University of Adelaide. Dr C. A. Morris (retired) was the project leader in NZ.

Beef & Lamb NZ Genetics funded the genotyping work and the analysis presented in this study.

REFERENCES

Adamsson, V., Cederholm, T., Vessby, B. and Riserus, U. (2014) *Food & Nutrition Res.* **58**:24114.

Bouwman, A.C., Bovenhuis, H., Visker, M.H.P.W. and Van Arendonk, J.A.M. (2011) *BMC Genetics* **12**:43.

Brooks, M.A., Choi, C.W., Lunt, D.K., Kawachi, H. and Smith, S.B. (2011) *J. Anim. Sci.* **89**(8):2556-2570.

Browning, S.R. and Browning, B.L. (2007) *Am. J. Hum. Genet.* **81**:1084-1097.

Costa, A.S., Silva, M.P., Alfaia, C.P., Pires, V.M., Fontes, C.M., Bessa, R.J. and Prates, J.A. (2013) *Lipids* **48**(4):369-381.

Gunstone, F.D., Harwood, J.L. and Padley, F.B. (1994) 'The Lipid handbook' Chapman & Hall, London.

Hill,W.G.,and Roberson,A. (1968) *Theor.Appl.Genet.* **38**:226–231.

Katan, M.B., Zock, P.L. and Mensink, R.P. (1994) *Am. J. Clin. Nutr.* **60**(suppl):1017-1022.

McVean, G. (2007) 'Handbookofstatistical genetics*'* 3rd ed. John Wiley and Sons, New York.

Morris, C.A., Bottema, C.D., Cullen, N.G., Hickey, S.M., Esmailizadeh, A.K., Siebert, B.D. and Pitchford, W.S. (2010) *Anim. Genet.* **41**:589-596.

Morris, C.A., Cullen, N.G., Glass, B.C., Hyndman, D.L., Manley, T.R., Hickey, S.M., McEwan, J.C., Pitchford, W.S., Botteman, C.D.K. and Lee, M.A.H. (2007) *Mammalian Genome* **18**:64-74.

Mozaffarian, D., Ascherio, A., Hu, F.B., Stampfer, M.J., Willett, W.C., Siscovick, D.S. and Rimm, E.B. (2005) *Circulation* **111**(2): 157-164.

Rincon, G., Islas-Trejo, A., Castillo, A.R., Bauman, D.E., German, B.J. and Medrano, J.F. (2012) *J. Dairy Res.* **79**(1):66-75.

Roy, R., Ordovas, L., Zaragoza, P., Romero, A., Moreno, C., Altarriba, J. and Rodellar, C. (2006) *Anim. Genet.* **37**(3):215-218.

Sargolzaei, M. (2010) Snppld v1.0, University of Guelph, Ontario, Canada.

Sargolzaei, M., Chesnais, J.P. and Schenkel, F.S. (2014) *BMC Genomics* **15**:478.

Tait, R.G., Zhang, S., Knight, T., Strohbehn, D. and Beitz, D.C. (2008) *Anim. Industry Report* AS 654, ASL R2285.

Yokota, S., Sugita, H., Ardiyanti, A., Shoji, N., Nakajima, H., Hosono, M., Otomo, Y., Suda, Y., Katoh, K. and Suzuki, K. (2012) *Anim. Genet.* **43**(6):790-792.

Youssef, M.K.E, Mostafa, B.M.D., Seliem, M.A.A. and Hashem, A.M.A. (2012) *Food and Public Health* **2**(2):34-39.

Zhang, S., Knight, T.J., Reecy, J.M. and Beitz, D.C. (2008) *Anim. Genet.* **39**(1):62-70.

Proc. Assoc. Advmt. Breed. Genet. 21: 57-60

EFFECTS OF *TEX11* AND *AR* POLYMORHISMS ON REPRODUCTION AND GROWTH TRAITS IN AUSTRALIAN BEEF CATTLE

L.T. Nguyen[1], G.M.F.D Camargo[2,], R.E. Lyons[3], S.A. Lehnert[4] and M.R.S. Fortes[1]

[1]University of Queensland, School of Chemistry and Molecular Biosciences, Qld 4072, Australia.
[2]Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal-SP, 14884-900, Brazil.
[3]University of Queensland, School of Veterinary Science, Gatton QLD 4343, Australia.
[4]CSIRO Agricultural flagship, Queensland Bioscience Precinct South, Qld 4067, Australia.

SUMMARY

Previous genome-wide association analyses indicated QTL regions located in X chromosome for scrotal circumference (SC) and percentage of normal sperm (PNS). The association between SNP in two potential candidate genes (*TEX11* and *AR*) on chromosome X and observed phenotypic variation of SC and PNS were analysed. As expected from QTL findings, these SNP could explain more than 1% of the additive genetic variance for SC. Three SNP in *TEX11* and a SNP in *AR* were significant for SC measurements taken at 12, 18 and 24 months of age. SNP in exon 1 of *TEX11* gene had extremely significant effects on SC12, SC18 and SC24 with *P*-values ranging from 10^{-39} to 10^{-46}. An association between a SNP in *TEX11* and weight measurements was also identified. Associations reported herein suggest that these SNP in *TEX11* and *AR* might aid genomic selection for SC and weight if included in genotyping panels.

INTRODUCTION

Fertility has important economic impact for livestock and fertility traits are considered in breeding programs. However, fertility is a rather complex phenotype that can be described by many indicator traits, such as scrotal circumference (SC) or sperm quality (Cammack *et al*. 2009). Traits considered as indicators of fertility are expressed late in life and mostly have low heritability (Cammack *et al*. 2009). Complexity, low heritability and late expression create challenges for selective breeding. Yet, some fertility traits are of moderate heritability, such as SC (heritability ranged from 0.29 to 0.78) (Cammack et al. 2009) and percentage of normal sperm (PNS, heritability = 0.35 (Kealey et al. 2006)). Bull fertility traits including SC and PNS are commonly measured at bull breeding soundness evaluation and can be used for improvement of fertility. Bull SC is also utilized for female fertility improvement due to its correlation with heifer age at puberty (Evans *et al*.,1999). Thus, SC and PNS are selection traits for beef cattle fertility.

Previous genome-wide association analyses reported QTL regions in chromosome X as associated with SC and PNS in Brahman and Tropical composite bulls (Fortes *et al*. 2012; Fortes *et al*. 2013). These QTL regions indicate that the *TEX11* and *AR* genes are candidates for identifying putative causative mutations. This study was carried out to identify and test putative causative mutations in these genes.

MATERIALS AND METHODS

Animals and phenotype. Animal care and Use Committee approval was not required for this research since samples and data used were from existing databases. Data were obtained from 1,178 Brahman bulls, 1,360 Tropical Composite bulls and 167 crossbreds (Tropical composite vs Brahman). These animals were the progeny of sires from Beef CRC. In total, 2705 bulls were analysed together in this study. Traits analysed were: SC and weight (WT) measured at 12, 18 and 24 months and PNS measured at 18 and 24 months. Measurement details for the Beef CRC

populations were described previously (Burns *et al*. 2013; Corbet *et al*. 2013). Genomes of 16 Brahman bulls (sires of genotyped animals) were utilized to generate VCF format files with variants information. Variant Effect Predictor (VEP) was used to predict the functional consequences of detected variants.

Genotyping and Linkage. Using Taqman assays, 2,705 bulls were genotyped for 4 SNP: Tex11-r38k, Tex11-g297d, Tex11-r696h and AR-intron6. Linkage disequilibrium (LD,r^2) was estimated pair-wise for genotyped SNP, using SVS software (Release 8.3.0, Golden Helix, Inc.).

Analyses. Association of selected SNP with SC, WT and PNS was examined using SVS software (Release 8.3.0, Golden Helix, Inc.) A mixed model analysis of variance was used to estimate the SNP effect and its significance level. The mixed model can be written as an equation: $Y_i = X\beta + Z\mu + S_j a_j + e_j$, where Y represents the phenotypic measurement for the ith animal, X is the incidence matrix relating fixed effects in β, Z is the incidence matrix relating to random additive polygenic effects of animal in μ, S is a vector of genotypes of each animal at SNP (j), a_j is the additive effect of the jth SNP, and e_j is the random residual effect. SNP is fitted was random and fixed effects were those of contemporary group (year, management group and breed). Age was fit as a covariant.

RESULTS AND DISCUSSION

Only 3 nsSNP discovered in *TEX11* and none in *AR* using the genome sequences available. These SNP are more likely to alter protein sequence and structure, and be beneficial or deleterious. It would however be inaccurate to state that all functional changes are based on protein coding sequence. The alteration of regulatory sites also can disrupt the expression of target genes (Knight *et al*. 2003). An intronic SNP in *AR* was also tested. From LD analysis, the SNP Tex11-r38k and Tex11-r696h were completely linked with r^2 value of 1. Therefore, it was impossible to differentiate the effects between these two SNP in all animals. Consequently, the SNP Tex11-r38k was used to represent both in the following results and discussion. For all the other pairs, the estimates of LD were lower than 0.6. The effect of these SNP could be interpreted separately.

The Tex11-g297d SNP had a slightly lower association with SC and WT measurements relative to the results obtained for Tex11-r38k, which had *P*-values in the range of 10^{-39} to 10^{-46} (Table 1). In Brito *et al*. (2002), increased SC has been related to increased sperm production but decreased semen quality. In our results, these same SNP were not associated with PNS, a measurement of semen quality. In 2003, Martínez-Velázquez reported that SC is positively correlated with growth traits. SNP that showed associations with both SC and WT could be expected. The SNP in *AR* showed an association with SC but not WT. We tested the relative relevance of Tex11-r38k to the QTL previously described (Fortes *et al*. 2012; Fortes *et al*. 2013) by fitting it as a fixed effect in the GWAS model. When Tex11-r38k was utilized as a fixed effect, the associations between common SNP in chromosome X (Illumina chip variants) and SC measurements were reduced (Figure 1). This result is consistent with expectations from causative variants that are able to explain the underpinning QTL.

Associations *TEX11* SNP with bull fertility were first studied by Lyons *et al*. (2013). The results obtained here validated that study. The substitution from G to A on *TEX11* might have negative effect on bull performance (smaller SC). Deleterious effect of Tex11-r38k was predicted according to SIFT. *TEX11* competes with estrogen receptor beta for a specific binding to HPIP protein (Yu *et al.* 2012). The *TEX11* protein region that binds to HPIP protein is between amino acids 370 and 947 (Yu *et al.* 2012), indicating that SNP Tex11-r696h that change amino acid at position 696 and are completely linked to SNP Tex11-r38k may be the best functional mutations. Further studies about biological role played by *TEX11* and its SNP in bull fertility are warranted.

Table 1.Significance and estimated effects of selected SNP on reproductive and growth traits in mixed bull population

Trait	SNP	p-value	Effect	SE	%Va
PNS18	Tex11-r38k	0.0124	-1.6716	0.6676	0.3180
	Tex11-g297d	0.1112	1.5053	0.9447	0.1291
	AR-intron6	0.1472	-1.4475	0.9982	0.1069
PNS24	Tex11-r38k	0.4681	-0.3611	0.4977	0.0215
	Tex11-g297d	0.2197	0.8396	0.6839	0.0614
	AR-intron6	0.7849	0.1848	0.6771	0.0030
SC12	Tex11-r38k	$2.39x10^{-39}$	-0.7431	0.0557	6.2944
	Tex11-g297d	0.0002	-0.2863	0.0775	0.5120
	AR-intron6	0.0105	-0.1937	0.0757	0.2467
SC18	Tex11-r38k	$6.03x10^{-46}$	-0.8270	0.0570	7.3867
	Tex11-g297d	0.0003	-0.2842	0.0794	0.4837
	AR-intron6	$1.04x10^{-05}$	-0.3430	0.0776	0.7339
SC24	Tex11-r38k	$2.44x10^{-43}$	-0.7708	0.0548	6.9741
	Tex11-g297d	0.0005	-0.2646	0.0765	0.4516
	AR-intron6	$1.94x10^{-05}$	-0.3200	0.0748	0.6900
WT12	Tex11-r38k	$7.04x10^{-08}$	-3.0673	0.5674	1.0926
	Tex11-g297d	0.0003	-2.7609	0.7704	0.4832
	AR-intron6	0.3354	-0.7250	0.7525	0.0351
WT18	Tex11-r38k	$3.97x10^{-08}$	-3.7138	0.6742	1.1329
	Tex11-g297d	0.0016	-2.8693	0.9110	0.3732
	AR-intron6	0.35401	0.8534	0.9206	0.0324
WT24	Tex11-r38k	$2.50x10^{-07}$	-3.7426	0.7237	1.0010
	Tex11-g297d	0.0007	-3.2871	0.9785	0.4249
	AR-intron6	0.1255	1.5132	0.9874	0.0887

CONCLUSION

Our results provide evidence for a key role of *TEX11* in male reproduction in beef cattle. The SNP Tex11-r38k and/or Tex11-r696h are proposed as functional mutations in *TEX11*. The *AR* gene remains as a candidate gene as its SNP was also associated with SC. As shown, SNP in these candidate genes influence SC and PNS in *Bos indicus* and their crossbreds. As a result, these associated SNP could be incorporated in low-density chips to facilitate genetic evaluation.

ACKNOWLEDGEMENTS

The authors thank the financial support of Meat and Livestock Australia (project B.NBP.0786). Fortes is supported by UQ Postdoctoral Fellowship. Nguyen is supported by UQ International Scholarship.

REFERENCES

Brito L.F., Silva A.E., Rodrigues L.H., Vieira F.V., Deragon L.A. and Kastelic, J.P. (2002) *Theriogenology,* 58: 1175.

Burns B.M., Corbet N.J., Corbet D.H., Crisp J.M., Venus B.K. *et al.* (2013) *Anim Prod Sci.,* 53**:** 87.

Cammack K.M., Thomas M.G. and Enns R.M. (2009) *Prof Anim Sci,* 25**:** 517.

Corbet N.J., Burns B.M., Johnston D.J., Wolcott M.L., Corbet D.H *et al.* (2013) *Anim Prod Sci,* 53**:** 101.

Evans J.L., Golden B.L., Bourdon R.M. and Long K.L.(1999) *J.Anim. Sci,* 77**:** 2621.

Fortes M.R.S., Reverter A., Kelly M., Mcculloch R. and Lehnert S.A. (2013) *Andrology,* 1**:**644.

Fortes M.R.S., Reverter A., Hawken R.J., Bolormaa S. and Lehnert S.A. (2012) *Biol Reprod,* 87**:** 58.
Kealey C.G., Macneil M.D.,Tess M.W., Geary T.W. & Bellows R.A. 2006. *J.Anim..Sci.* 84: 283.
Knight J.C., Keating B.J., Rockett K.A. and Kwiatkowski D.P. (2003) *Nat Genet,* 33**:** 469.
Lyons R.E., Loan, N.T., Dierens L., Fortes M.R., Kelly M., Mcwilliam S.S., *et al*. (2014) *BMC Genet,* 15**:** 6.
Martinez-velazquez G., Gregory K.E., Bennett G.L. and Van vleck L.D. (2003) *J.Anim.Sci,* 81: 395.
Parkinson T. J. (2004) *Vet J*, 168: 215.
Patnala R., Clements J. and Batra J. (2013) *BMC Genet*, **14**, 39
Yu Y.H., Siao F.P., Hsu L.C.L. and Yen P.J. (2012) *Mol Endocrinol*. 26: 1669

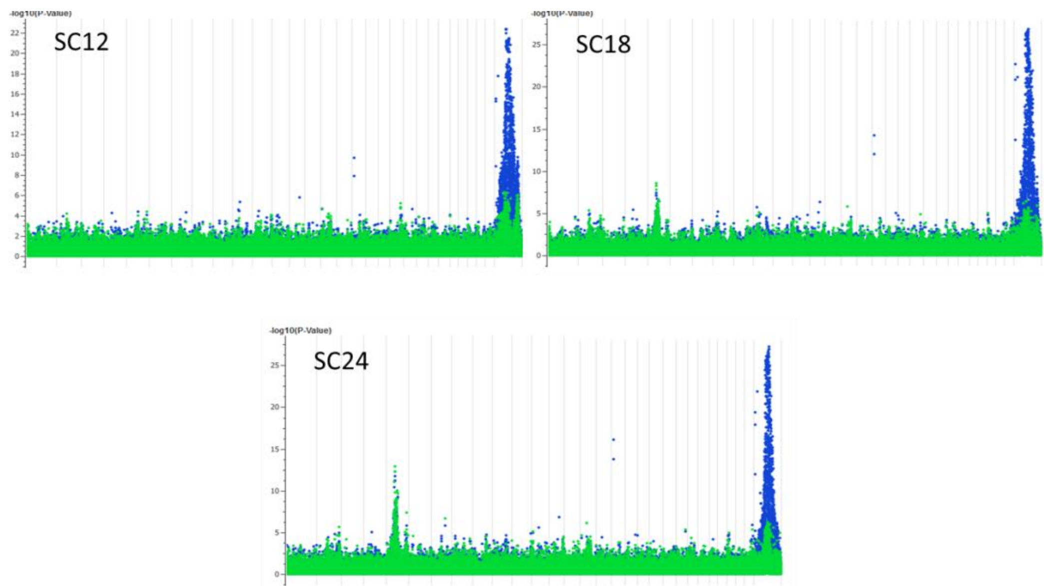


Figure 1. The association between SNP in 30 chromosome regions of beef cattle and scrotal circumference at 3 different ages. The chromosomal positions are in the x-axis and –Log (*P*-values) are in the y-axis. The blue indicates effect of all SNP on SC, the green reveals effect of SNP on SC with Tex11-r38k as a fixed effect.

Proc. Assoc. Advmt. Breed. Genet. 21: 61-64

USING PROTEIN QTL TO DISENTANGLE VARIANTS EFFECTING PROTEIN PERCENTAGE IN MILK NEAR THE CASEIN COMPLEX

K.E. Kemper[1], M.J. Carrick[2] and M.E. Goddard[1,3]

[1] Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Victoria, Australia
[2] Berghan Carrick Consulting, Moonee Ponds, Victoria, Australia
[3] Dept. of Economic Dev., Jobs, Transport and Resources, AgriBio, Bundoora, Victoria, Australia

SUMMARY

Casein proteins comprise about 80% of the protein in bovine milk. The casein complex occupies a 300 kb region of the genome on BTA6. To disentangle the number of variants affecting protein percentage in bovine milk near the casein complex, we use single variant regressions with imputed full sequence variants and quantified α_{S1}-, β- and κ-casein protein levels in 444 Holstein Friesian cows. We find 2 variants, located near CSN3 (coding κ-casein) and within CSN1S1 (which codes for α_{S1}-casein), with independent effects on P% and which affect concentration of their corresponding casein gene products. Previously described protein polymorphisms in the casein genes were sometimes associated with the quantity of their respective proteins but it seems unlikely that these variants are causing variation in casein concentrations.

INTRODUCTION

Four types of casein proteins (α_{S1}-, α_{S2}-, β- and κ-casein) constitute about 80% of the protein in milk and the genes encoding the casein proteins are located in a 300kb region on *Bos taurus* autosome (BTA) 6 (Table 1). Polymorphisms in the amino acid sequence of these proteins have been known for many years and have been found to be associated with milk protein yield and concentration. However, the associations have not been consistent across studies perhaps because the mutations causing variation in amount of casein are not the same as those causing differences in amino acid sequence although they may be in linkage disequilibrium with them (reviewed by Goddard & Wiggans 1999).

Table 1. Genomic location of the casein genes on *Bos taurus* autosome (BTA) 6[*]

Gene description	Symbol	location (bp)	protein product
Bos taurus casein alpha-S1, mRNA (+)	CSN1S1	87,141,556-87,159,096	α_{S1}-casein
Bos taurus casein beta, mRNA (-)	CSN2	87,179,502-87,188,025	β-casein
Bos taurus casein alpha-S2, mRNA (+)	CSN1S2	87,262,457-87,280,936	α_{S2}-casein
Bos taurus casein kappa, mRNA (+)	CSN3	87,378,398-87,392,750	κ-casein

[*]Other genes also located in the region. Locations from UMD3.1 (www.ensembl.org/Bos_taurus/). The forward (+) or reverse (-) orientations for transcription are indicated after the gene description.

Kemper *et al.* (2015) identified a sequence variant (Chr6:87296809) as affecting protein content (P%) from a multi-trait meta analysis of Holstein and Jersey cattle. This variant was located within an intergenic region, closest to the *Bos taurus* casein alpha-S2 coding region (CSNS2). However, this analysis did not exclude the possibility that this variant was associated with cumulative effects of several different underlying P% causal variants in the region. The simplest hypothesis is that mutations in the regulatory region of each casein gene cause variation in the amount of that casein produced and therefore in the amount of total protein. The aim of this study is to disentangle the P% QTL observed in Holstein cattle near the casein complex by using phenotypes consisting of P% and quantification of three casein proteins (α_{S1}-, β- and κ-casein).

MATERIALS AND METHODS

Overview. The paper aims to identify variants (either causal variants or variants in strong LD with the causal variant) underlying α_{S1}-, β- and κ-casein concentrations in a small dataset and then to test whether or not these variants can explain the variation in P% due to the QTL near the casein complex. The methods are detailed below and consist of single variant regression for α_{S1}-, β- and κ-casein concentration, followed by conditional single regression analysis for P% using imputed whole genome sequence data.

Phenotypes and genotypes. There are two datasets. The first dataset consists of genotypes and phenotypes for 444 cows measured for α_{S1}-, β- and κ-casein concentration (mg/g) using capillary zone electrophoresis (Kanning, Casella & Oliman 1993) on combined morning and afternoon milking at two sampling days, approximately 6 weeks apart. A model with fixed (mean concentration, breed, 4th order polynomials for age & days-in-milk) and random (herd, permanent environment (PE), animal) effects was fitted to the data and trait-deviations for animals constructed as the average of PE, animal and residual effects for animals with two measurements. Genotypes were available for the 50K bovine single nucleotide polymorphism (SNP) chip and these genotypes were imputed to the high-density array (632,002 SNP) following Erbe *et al.* (2012). Protein types for α_{S1}-, β- and κ-casein were determined using gel electrophoresis following Ng-Kwai *et al.* (1984).

The second dataset, described by Kemper *et al.* (2015), consists of P% phenotypes and genotypes of 632,002 (real and imputed) high-density SNP for 8478 Holstein cows. Unlike Kemper *et al.* (2015), this analysis uses only Holstein animals.

Sequence variants. Sequence variants consisted of SNP and small INDEL from a 5 Mb region centred on the casein complex (BTA6: 84.5-89.5 Mb). Data were obtained from run 4 of the 1000 bull genomes project (Daetwyler *et al.* 2014).

Imputation and the association study. Sequence variants were imputed into the 2 datasets for the target region on BTA6 using Minimac (Fruchberger, Abecasis and Hinds 2015) and 260 sequenced Holstein animals as the reference population. The association study for each phenotype used EMMAX (Kang et al. 2010) following Kemper *et al.* (2015). Multi-allelic protein polymorphisms were treated as a series of contrasts (i.e. A1 & B types vs. A2 for β-casein). The conditional analysis for P% was also conducted using EMMAX.

RESULTS AND DISCUSSION

Variants associated with individual casein concentrations. The most significant results were obtained for κ-casein, followed by β- and α_{S1}-casein concentrations (Table 2). The variant most highly associated with κ-casein concentration was Chr6:87405588, located about 13 kb downstream of CSN3 ($P = 7.7 \times 10^{-12}$). Similarly for β- and α_{S1}-casein, the most significant variants were outside the coding regions for the genes, where Chr6:87098077 is 43 kb upstream of CSN1S1 and Chr6:87206907 is 19 kb upstream of CSN2 (N.B. that CSN2 is transcribed on the reverse strand). However, there are a number of other sequence variants which are also highly associated with the casein concentrations and any one of these could be the causal mutation (Figure 1).

Variants associated with protein variants. The frequency of the protein polymorphisms varied widely between the casein genes, with β-casein A1 and A2 variants being of intermediate frequency (0.44 & 0.51 respectively), the κ-casein B variant having a relatively high frequency compared to the C variant (0.78 vs. 0.22) and the β-casein B having a low frequency (0.04). There were few observations of the C allele for α_{S1}-casein and the A3 allele at β-casein, effectively rendering the α_{S1}-casein protein type monomorphic.

The missense mutations causing the known protein polymorphisms were associated with the protein variants as expected. In each case, there were a number of other sequence variants that

also associated with the protein type due to the high degree of LD in the region and the small sample size (Table 2). These sequence variants were 20-30 kb away from the variants identified as influencing the quantity of these proteins. Both the κ-casein B/C protein polymorphism (Chr6:87390576; P = $4.3x10^{-11}$) and the β-casein B protein polymorphism (Chr6:87181453; P = $1.3x10^{-4}$) were strongly associated with the concentrations of their respective proteins. However, these variants were more than 1 $\log_{10}$ unit from the most significant variant and it seems likely that they are in LD with variants affecting the protein concentrations.

Table 2. Most significant sequence variants for casein concentrations and protein types, where the variants within 1 $\log_{10}$ unit (number; location range, bp) assesses level of confidence in the top variant

phenotype	top variant (P value)	additional variants within 1 $\log_{10}$ unit
κ-casein conc.	Chr6:87405588 ($7.7x10^{-12}$)	132 (87,333,107 – 87,407,175)
α_{S1}-casein conc.	Chr6:87098077 ($9.2x10^{-6}$)	18 (87,085,525 – 87,154,594)
β-casein conc.	Chr6:87206907 ($8.3x10^{-6}$)	4 (87,090,414 – 87,115,771)
κ-casein B vs. C protein type[1]	Chr6:87393434 ($4.7x10^{-14}$)	122 (87,363,855 – 87,405,868)
β-casein A1/B vs. A2 protein[2]	Chr6:87184548 ($2.2x10^{-167}$)	8 (87,169,673 – 87,184,548)
β-casein A1/A2 vs. B protein[3]	Chr6:87186827 ($1.8x10^{-319}$)	2 (87,185,552 – 87,189,903)

[1]Chr6:87390576 is the mutation causing Ile>Thr (ref>alt) substitution in κ-casein and was ranked 29th in the analysis; [2]Chr6:87181619 is the mutation causing the His>Pro substitution in β-casein and was ranked 5th; [3]Chr6:87181453 is the mutation causing the Ser>Arg substitution in β-casein and was ranked 3rd.

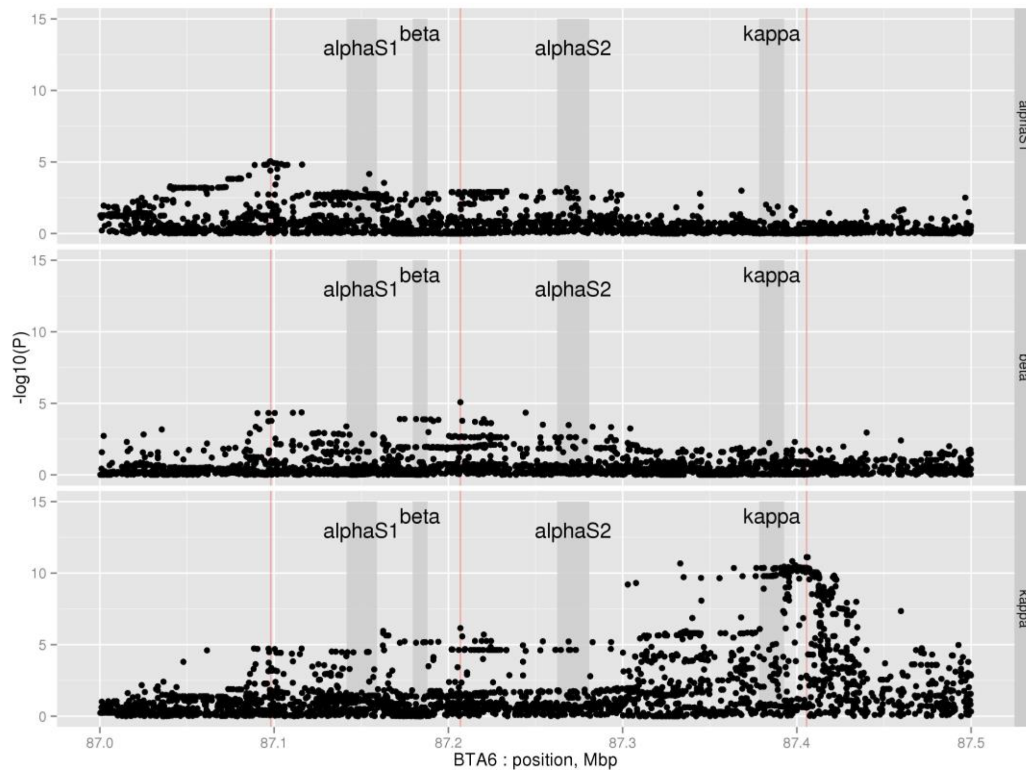


Figure 1. Association study in the casein region for quantity of α_{S1}-, β-, and κ-casein. The most significant variant for each trait (Chr6:87098077, Chr6:87206907 & Chr6:87405588) is highlighted with vertical lines.

Dissecting the P% QTL observed in the casein region. The variant identified by Kemper *et al.* (2015), Chr6:87296809, was highly significant for P% ($6.7x10^{-14}$) but was not within a $\log_{10}$ unit of the top variant for any of the three individual casein concentrations. Thus the analyses aimed to discover independent variants for P%, based on the most significant κ-casein variant, followed by any variants remaining significant ($P < 1x10^{-6}$) after a conditional analysis on the κ-casein variant. It was found that fitting two variants (Chr6:87405588 and Chr6:87154594) reduced all other variants to $P > 1x10^{-6}$ (Figure 2). The Chr6:87154594 variant was located within an intron of CSN1S1 and was within 1 $\log_{10}$ unit of the top variant for α_{S1}-casein concentration ($P = 6.8x10^{-5}$). The most significant variant for β-casein concentration (Chr6:87206907) was not significantly associated with P% (P = 0.82) after adjusting for Chr6:87405588. This suggests there is not an independent effect of β-casein concentration on P%. It is possible that Chr6:87206907 is capturing the effect of a haplotype which affects both κ- and β-casein concentrations.

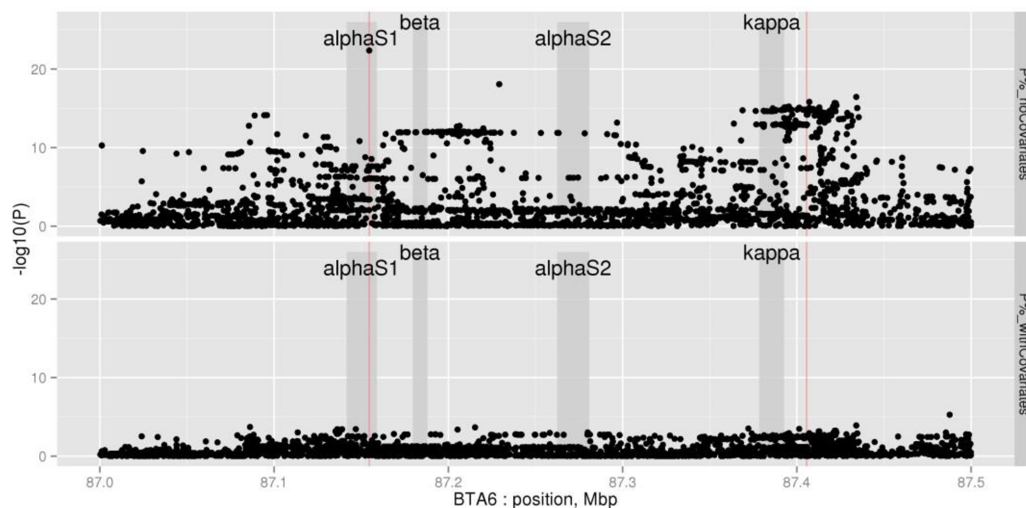


Figure 2. Association study in the casein region for P% without fitting covariates (top), and after fitting covariates of Chr6:87405588 and Chr6:87154594 (bottom, variants indicated by vertical lines).

We conclude that the significant result for Chr6:87296809 (Kemper *et al.* 2015) was likely due to the cumulative effects of at least two variants affecting P% in the casein complex. Our results suggest 2 independent variants that influence κ- and α_{S1}-casein concentrations and therefore cause variation in P%. It seems that these variants are distinct (but sometimes associated with) the known mutations causing the protein polymorphisms for α_{S1}-, κ- and β-casein.

REFERENCES

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., *et al.* (2014) *Nat Genet* **46:**858.

Erbe, M., Hayes, B.J., Matukumalli, L.K., Gosawami, S., *et al.* (2012) *J Dairy Sci* **95:**4114.

Fruchberger, C., Abecasis, G.R., Hinds, D.A. (2015) *Bioinformatics* **31:**782.

Goddard, M.E. and Wiggans, G.R. (1999) In 'The genetics of cattle', pp. 522-525, editor R. Fries & A. Ruvinsky, CABI Publishing.

Kang, H.M., Sul J.H., Service S.K., Zaitlen N.A, *et al.* (2010) *Nat Genet* **42:**348.

Kanning, M., Casella, M. and Olieman, C. (1993) *LC-GC Int* **6:**701.

Kemper K.E., Hayes B.J., Daetwyler, H.D. & Goddard M.E. (2015) *J Anim Breed Genet.* **132:**121.

Ng-Kwai-Hang, K.F., Hayes, J.F., Moxley, J.E. and Monardes, H.G. (1984) *J Dairy Sci* **67:**835.

Proc. Assoc. Advmt. Breed. Genet. 21: 65-68

A COMPARISON OF GENOTYPES GENERATED BY INFINIUM BEADCHIPS AND A TARGETED GENOTYPE BY SEQUENCING METHOD IN CATTLE

A.J. Chamberlain[1], P. Beatson[4], B.A. Mason[1], C. Prowse-Wilkins[1], C.M. Reich[1], C. Schrooten[5], B.J. Hayes[1,2], J. Tibbits[1,3], and M. Hayden[1]

[1] Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia; [2] La Trobe University, Bundoora, Victoria, Australia; [3] Melbourne University, Victoria, Australia; [4]CRV, Hamilton, New Zealand; [5]CRV, 6800 AL, Arnhem, the Netherlands

SUMMARY

The cost of SNP genotyping is of major importance in genomic selection programs. One possibility for reducing the expense of applying genomic selection would be to take advantage of recent dramatic decreases in the cost of genome sequencing, by using genotyping-by-sequencing (GBS) techniques to provide cheaper genotypes. This paper presents a GBS method that can target individual variants and therefore any SNP of interest. Comparing array and GBS genotypes from 471 individuals for 5119 SNP we show that with GBS we can achieve sample call rates of 93%, as compared with 95% for arrays, and that genotypes called from the GBS are 98% concordant with those from SNP arrays. With further refinement of the custom reference we will be able to achieve higher call rates and genotyping accuracy.

INTRODUCTION

Genomic selection has now been widely implemented in dairy cattle industries worldwide, predominantly for the selection of sires. Genomic selection uses SNP genotypes to estimate SNP effects in a reference population, such reference populations have been established within most dairy countries. These SNP effects are used along with SNP genotypes of individual animals to make genomic predictions of breeding value in selection populations. Hence the cost of SNP genotyping is of major importance in genomic selection programs. Array based genotyping methods have been widely used to genotype cattle, particularly the Illumina Infinium 7K, 50K and HD beadchips. These are robust and accurate genotyping platforms, however the costs required to genotype an animal can be too high, preventing many dairy farmers from using such tests for heifer selection for example (Pryce, Hayes 2012).

Next generation sequencing has dramatically decreased in cost over the past decade and so many have turned to GBS techniques to provide cheaper genotypes for genomic selection programs. Many GBS techniques rely on the cleavage of DNA with restriction enzymes to generate a pool of DNA fragments, which are sequenced to enable SNP discovery and genotype calling (Elshire *et al.* 2011). These methods provide SNP that are randomly located across the genome. However much work has now been done to identify the most informative variants to use for genomic selection, and this work will continue into the future. Therefore, GBS techniques that can target individual variants would be more informative and flexible than those that use random variants. The GBS technique presented here is one such method. In this paper we present results demonstrating the performance and accuracy of this targeted GBS technique on 479 Holstein cows for 9102 SNP.

MATERIALS AND METHODS

Probes were designed to the flanking sequencing of 9102 target SNP, 5119 of which form part of the Illumina Infinium BovineSNP50 beadchip. These probes were used to capture DNA fragments containing the target sites from 479 bovine DNA samples in a method similar to that of (Shen *et al.* 2013). The products were PCR amplified using indexed primers that provided

compatibility with sequencing on the HiSeq2000 genome analyser platform, and sequenced using single read chemistry.

An informatics pipeline was created to perform the following steps: 1) Sequence reads were trimmed of adapter sequence and poor quality bases (qscore < 20) using in-house scripts. 2) The quality filtered reads were aligned using BWA v0.7.7 (Li, Durbin 2009) to a custom reference genome (described below) allowing 4 mismatches and performing an exhaustive search for each read. 3) Samtools v0.1.19 (Li *et al.* 2009) mpileup tool with an input file listing all target SNP sites was used to create vcf files for all samples which in turn were used to create allele counts at all 9102 target SNP sites. 4) Allele counts were used to call genotypes, where the total count must be 6 or greater and a heterozygote had to have a minimum minor allele frequency of 0.167 (1 in 6 counts). Where the total allele count was <6, the genotype was set to NC (no call). 5) The genotypes (in UMD3.1 forward format) were then converted to TOP-TOP format (http://www.illumina.com/documents/products/technotes/technote_topbot.pdf).

471 of the DNA samples were also genotyped with the Illumina Infinium BovineSNP50 beadchip as per manufacturer's instructions, with genotype calls output in TOP-TOP format.

Starting contigs, consisting of the SNP and it's flanking sequence captured by the probes, were created for each SNP. Additional SNP within those sequences, discovered in the 1000 bull genomes (Daetwyler *et al.* 2014) run3.0 dataset, along with phased genotypes of all Holstein and Jersey animals, were used to create known haplotypes for each contig. Starting contigs were then edited to reflect the new haplotypes and new contigs created. All contigs were then combined into the custom reference genome which was used in the above informatics pipeline.

RESULTS AND DISCUSSION

Custom reference. The custom reference consisted of 27,918 contigs for 9102 target SNP. Target SNP had up to 17 SNP within the flanking sequence, and up to 53 different haplotypes per target. For targets with such large numbers of flanking SNP and so many different haplotypes, using standard reference genomes to align reads would result in reduced numbers of reads mapping as many reads would exceed the 4 allowed mismatches. This in turn would result in inaccurate genotype calls. Therefore custom reference building is essential, and must be revised periodically to incorporate new haplotypes within the population being genotyped.

GBS performance. On average 1.9 million reads were generated per library, of which 1.2 million reads passed quality filtering and trimming. On average 82% of all reads for each sample mapped to the custom reference genome. The rank ordered distribution of log10(read counts) between samples (Figure 1a) and between SNP across samples (Figure 1b) showed each sample had relatively uniform representation and that the majority of SNP were evenly covered. The method is very flexible and so the assay can be changed to include different or additional SNP. Therefore in the future SNP that failed, ie had very low read count, would be removed from the assay and new SNP that are deemed important would be included. This dataset presented here targets 9102 SNP, however this could be increased to very large numbers if need be, however proportionally more sequence reads would be required. The informatics pipeline for each sample took on average 4 hours to run, using less than 2 gigabases of memory.

Table 1. Average call rate for GBS and Infinium genotypes and concordance between the two for SNP and samples, where N is the number of SNP or samples and SD is the standard deviation

	N	GBS call rate (SD)	Infinium call rate (SD)	Concordance (SD)
SNP	5119	92.51% (0.175)	99.96% (0.002)	97.40% (0.069)
Sample	471	92.51% (0.073)	95.05% (0.067)	97.74% (0.047)

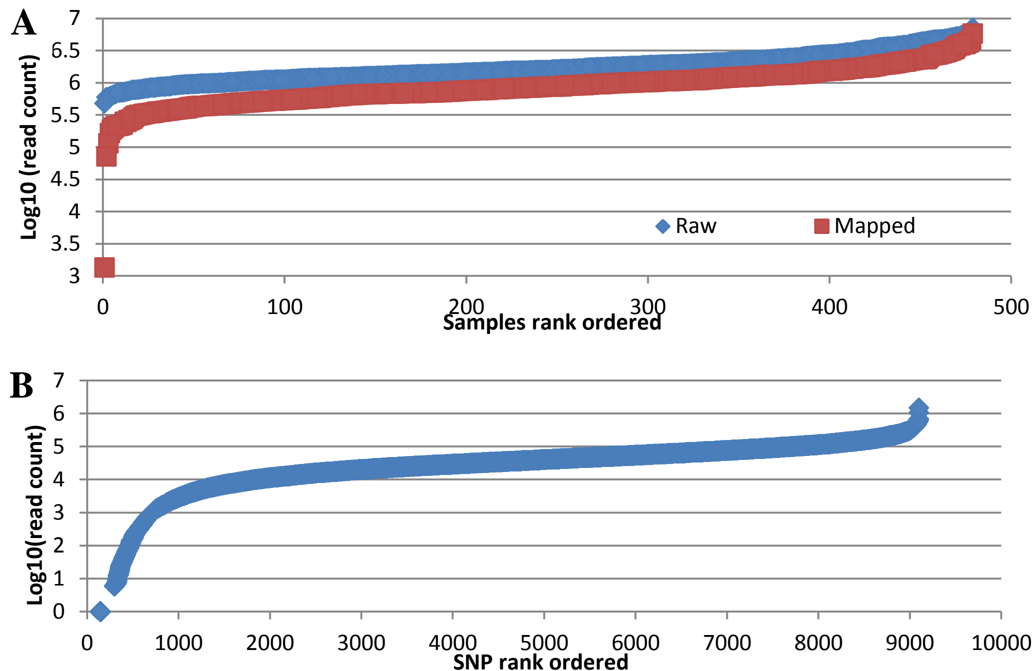


Figure 1. Distribution of reads amongst A) 479 samples (rank ordered) for raw and mapped reads and B) 9102 SNP (rank ordered) for mapped reads only. Read counts are expressed as $\log_{10}$(read count).

GBS and Infinium Beadchip concordance. We calculated the call rate and concordance for the targeted GBS and infinium genotypes for each sample for the 5,119 SNP common to both assays (Table 1). The average call rate for the 471 samples was 93% and 95% for the targeted GBS and infinium assays, respectively. The concordance between assays was 98% (Table 1). The majority of samples had both high call rates and high concordance (Figure 2a). Only one sample failed to generate genotype calls in the targeted GBS assay. In a separate experiment, we observed a correlation between the proportion of reads mapping and the amount of DNA used in the targeted GBS assay. This experiment showed a consistent high proportion of reads mapped could be achieved with an input amount of >400 ng DNA (Figure 3). Where samples have high proportion of reads mapped call rates are maximised.

While the average genotype concordance between the targeted GBS and Infinium assays was high (97%, Table 1), there was a subset of SNP with low concordance (Figure 2b). Upon closer investigation it was found that the target region for those SNP had additional variants, either other SNP or indels, that were previously unknown. As these variants were not represented in our custom reference, they often caused the number of mismatches to exceed the limit of 4 specified in the alignment. This resulted in not all of the reads associated with the target loci being mapped, and therefore incorrect genotype calling. Further work is currently being undertaken to discover new variants to update the custom genome and improve genotype calling accuracy.

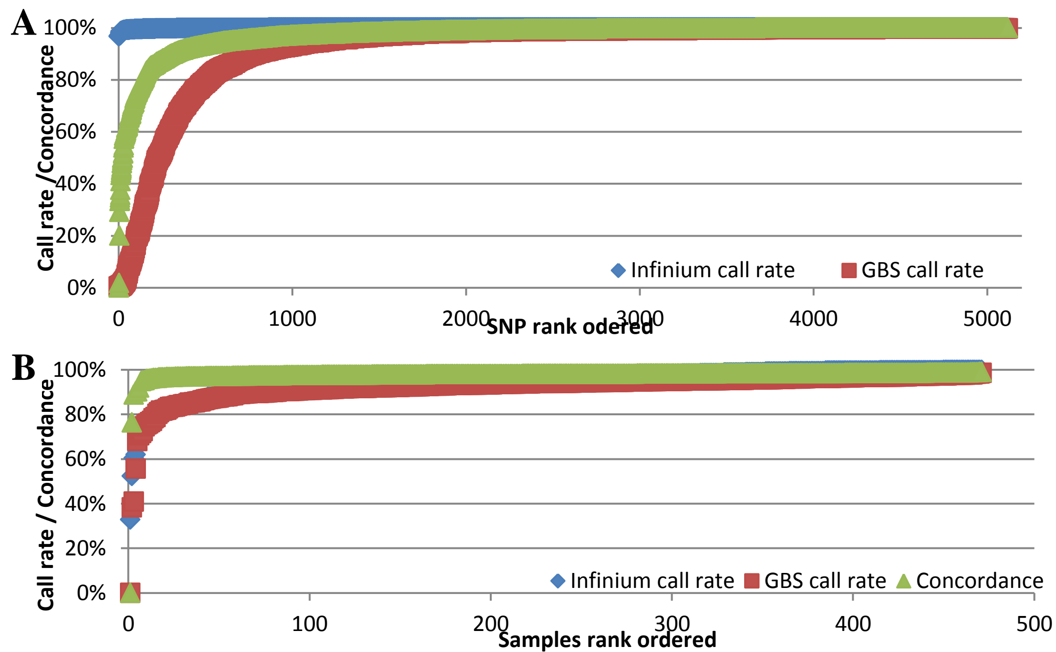


Figure 2. Infinium and GBS call rates as well as concordance between the two genotypes for A) samples (rank ordered) and B) SNP (rank ordered).

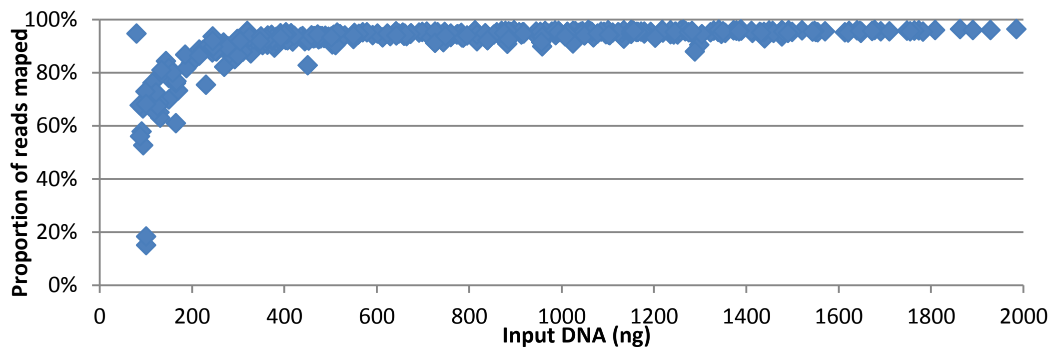


Figure 3. Relationship between the amount of input DNA in nanograms and the proportion of reads mapped to the custom genome for the GBS assay.

REFERENCES

Daetwyler, H.D., Capitan, A., Pausch, H., *et al.* (2014) Nature Genet. **46**:858.
Elshire, R.J., Glaubitz, J.C., Sun, Q., *et al.* (2011) *PLoS One.* **6**:e19379.
Li, H., Durbin, R. (2009) Bioinformatics **25**:1754.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J. *et al.* (2009) Bioinformatics **25**:2078.
Pryce, J., Hayes, B. (2012) Animal Production Science **52**:180.
Shen, P., Wang, W., Chi, A.-K., Fan, Y., Davis, R., Scharfe, C. (2013) Genome Medicine **5**:50.

Proc. Assoc. Advmt. Breed. Genet. 21: 69-72

IDENTIFICATION OF LOCI ASSOCIATED WITH PARASITE RESISTANCE IN AUSTRALIAN SHEEP

M. Al Kalaldeh, J.H.J. van der Werf and C. Gondro

University of New England, School of Environmental and Rural Science, Armidale 2350, Australia

SUMMARY

This study aimed to identify loci underlying variation in parasite resistance, as measured by worm egg count (WEC), in a large multi-breed sheep population using genome-wide association studies (GWAS) and regional heritability mapping (RHM) approaches. A total of 7153 animals with both genotype data and WEC phenotypes were included in this analysis. Strong evidence of association was observed on chromosome 2 by both approaches. However, RHM had a greater power to identify loci than GWAS analysis. RHM identified an additional region at the genome-wide significance level on chromosome 6. This region was also previously found to be associated with mastitis resistance and facial eczema susceptibility in sheep, indicating that some pleiotropic effects are possibly affecting a wide range of sheep diseases. Three other regions on chromosome 1, 3 and 24 reached the suggestive threshold. However, the regions accounted for a small proportion of genetic variance ($h_g^2 < 0.01$). It seems that parasite resistance is a complex disease with a large number of genes involved in the mechanism of resistance.

INTRODUCTION

Gastrointestinal nematode infections are one of the most important health problems affecting sheep and other grazing ruminants in Australia and worldwide. Selection for parasite resistance has been suggested as a viable method for parasite control (Roeber *et al.*, 2013). Most breeding programs for parasite resistance are based on phenotypic indicators, particularly worm egg counts (WEC) in faeces, but trait measurement is unattractive, costly and time consuming. Therefore, it would be very useful to select directly for parasite resistance. To date, several quantitative trait loci (QTL) mapping studies have been conducted for parasite resistance in sheep (e.g. Dominik *et al.*, 2010 and Marshall *et al.*, 2009). However, little overall consensus has emerged from these studies. This may be due to the physiological complexity of parasite resistance, and the fact that these studies are very diverse, involving a variety of analytical approaches, experimental designs, parasite species and sheep breeds. Further, genome-wide association studies (GWAS) for complex diseases, such as parasite resistance, have generally failed to explain the majority of genetic variation influencing the trait (Kemper *et al*. 2011). The objective of this study was to identify loci underling variation in parasite resistance in a multi-breed sheep population.

MATERIALS AND METHODS

Animals. Parasite resistance trait, as measured by WEC, was investigated in a multi-breed sheep population from the Sheep Cooperative Research Centre information nucleus flock (INF). A total of 7,539 animals with both genotype data and WEC phenotypes were included in this analysis. Various breeds were represented in the population (Table 1) but with a significant proportion of Merino sheep, and only this breed had a substantial proportion of purebred animals. The remaining breeds were mainly represented by their crosses with Merino (van der Werf *et al*. 2010).

Genotypes. Animals were genotyped using the 50k Ovine marker panel (Illumina Inc., SanDiego, CA, USA). SNPs were removed if they had a minor allele frequency (MAF) < 1%, an Illumina Gentrain score (GC) less than 0.6, a call rate less than 95%, or not in Hardy-Weinberg equilibrium. Furthermore, positions of SNPs were obtained from the latest sheep genome

Ovis_aries_v3.1, and any SNP with unknown position was removed. After applying these quality measures, 7,539 animals and 48198 SNPs were retained.

Table 1. Proportions of different breeds in the population

Breed	BL	COR	COOP	EF	WD	PD	TEX	AF	PS	MER
Proportion (%)	11.1	0.8	10	0.7	0.4	1.8	2.3	2	1.1	69.8

Border Leicester: **BL**, Corriedale: **COR**, Coopworth: **COOP**, East Friesian: **EF**, White Dorper:**WD**, Poll Dorset: **PD**, Texel: **TEX**, Australian Finnsheep: **AF**, Prime Samm: **PS**, Merino:**MER**

Data analysis. Genome-wide association studies (GWAS) and regional heritability mapping (RHM) approaches were performed using ASReml-R (Butler *et al.,* 2009). GWAS was performed using the GRAMMAR approach (Aulchenko *et al.,* 2007). In the first step, we fitted the following animal model to the data:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Z_1 q} + \boldsymbol{Z_2 a} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is a vector of cube root transformed WEC records, $\boldsymbol{X}$ is a design matrix of fixed effects, $\boldsymbol{b}$ is a vector of fixed effects, $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ are design matrices of random effects, $\boldsymbol{q}$ is a vector of random breed effects, $\boldsymbol{a}$ is a vector of random genetic effects, and $\boldsymbol{e}$ is the vector of residuals. The following distributions were assumed: $\boldsymbol{q} \sim N(0, I\sigma_q^2)$, $\boldsymbol{a} \sim N(0, A\sigma_a^2)$ and $\boldsymbol{e} \sim N(0, I\sigma_e^2)$, where $\boldsymbol{A}$ is the numerator relationship matrix (NRM) calculated from deep pedigree records, $\boldsymbol{\sigma_a^2}$ is the additive genetic variance explained by pedigree, $\boldsymbol{\sigma_q^2}$ is the variance of breed effects, and $\boldsymbol{\sigma_e^2}$ is the residual variance. The fixed effects were sex, rearing type × birth type, contemporary group (flock site × group of management × year of birth), age of animal at WEC recording and its quadratic polynomial. Second, residuals obtained from the animal model were treated as corrected phenotypes for a single- SNP regression:

$$\boldsymbol{\hat{y}} = \boldsymbol{1\mu} + \boldsymbol{Wa} + \boldsymbol{e}$$

where $\boldsymbol{\hat{y}}$ is a vector of adjusted phenotypes, $\boldsymbol{\mu}$ is the overall mean, $\boldsymbol{W}$ is a vector of a single SNP`s genotype for each of the animals, $\boldsymbol{a}$ is the effect size of the SNP, and $\boldsymbol{e}$ is the vector of residuals. The second approach was RHM, in which each chromosome was divided into windows of predefined number of SNPs, and the variance attributable to each window was calculated. In this analysis, two window sizes were used, 100-SNP and 50-SNP windows. The following model was fitted to the data:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Z_1 q} + \boldsymbol{Z_2 a} + \boldsymbol{Z_3 g} + \boldsymbol{e}$$

where the terms are as described in the animal model, and $\boldsymbol{g}$ is the regional genomic effect estimated from SNPs within each window. $\boldsymbol{g}$ was assumed to be distributed as $N(0, G\sigma_g^2)$, where G is the regional genomic relationship matrix built from SNPs within each window, and σ_g^2 is the regional genomic variance. Phenotypic variance, σ_p^2, was then given by $\sigma_q^2 + \sigma_a^2 + \sigma_g^2 + \sigma_e^2$. The whole heritability was calculated as $\mathrm{h_a^2} = \sigma_a^2/\sigma_p^2$, whereas the regional heritability was calculated as $\mathrm{h_g^2} = \sigma_g^2/\sigma_p^2$. Significance thresholds of GWAS and RHM were determined using the Bonferroni correction (significance threshold = α / N, where N is the number of tests) at the genome-wide (α =0.05) and suggestive (α=1) levels.

RESULTS AND DISCUSSION

The most significant GWAS results were observed on chromosome 2. Two SNPs, OAR2_119123707.1 and OAR2_119557086.1, were significantly associated with parasite resistance at the genome-wide and suggestive levels, respectively (Figure1). These results were also confirmed with RHM using 100 SNP window size (Figure 2). Both of GWAS and RHM

analyses generally agreed when there was a strong evidence of association (e.g.: chromosome 2). In this study, however, RHM detected more genomic regions significantly associated with parasite resistance therefore suggesting the method has greater power than GWAS analysis. For example, RHM using 100 SNP window size identified a significant region on chromosome 6 that was below the suggestive level by GWAS analysis. Furthermore, a region on chromosome 24 reaching the suggestive threshold was detected only with RHM. However, all significant regions identified by RHM explained a small proportion of WEC variation (RHM ranged from 0.0036 to 0.01), indicating that parasite resistance is a largely polygenic trait with a large number of loci involved in conferring resistance. Nagamine *et al.* (2012) showed that RHM captured more of the genetic variation than a single-SNP GWAS approach, especially when associated SNPs have very small effects to be declared significant at the genome wide level.

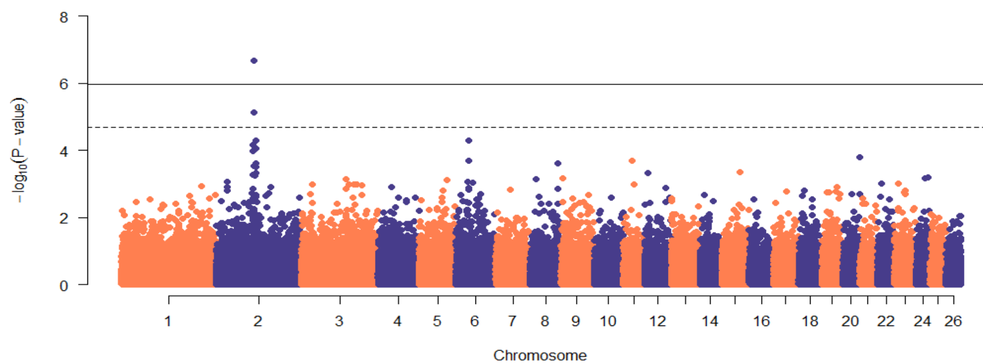


Figure1. Manhattan plot of GWAS results.The solid line represents the genome-wide significance threshold (α = 0.05) and the dashed line represents the suggestive threshold (α = 1).

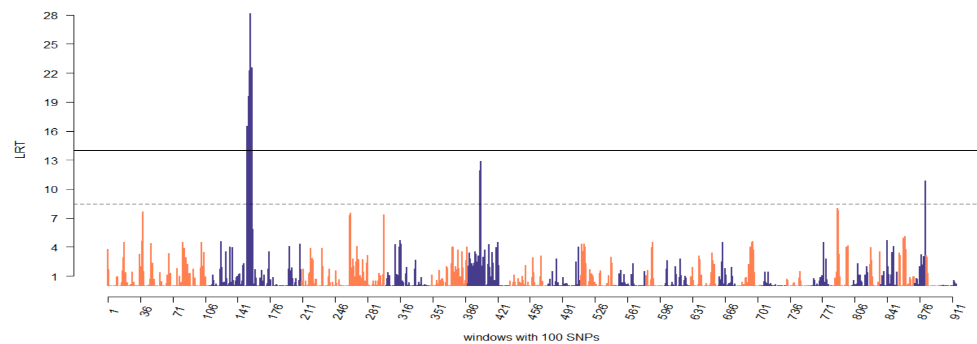


Figure2. Regional heritability mapping (RHM) across the genome. The solid line represents the genome-wide significance threshold (α = 0.05) and the dashed line represents the suggestive threshold (α = 1).

Significant regions with 100 SNP window size as well as those below the suggestive level were also analysed with a 50 SNP window size. The 50 SNP window size analysis confirmed the significant regions on chromosome 2 and 6, but did not confirm the region on chromosome 24. From all the region below the suggestive level, only two regions on chromosome 1 and 3 were significant. The results for RHM using 50-SNP window size are given in Table 2. Comparison

with other studies showed that significant region in chromosome 2 was contained within previously identified QTLs for parasite resistance (Hu *et al.,* 2013). This region has also been found to be associated with mastitis resistance in sheep (Jonas *et al.* 2011). Candidate genes in this region include: DEAD box polypeptide 60 (DDX60) and annexin A10 (ANXA10), which their expression found to be involved with immune response. Significant region in chromosome 6 has recently been identified by Riggio *et al.,* (2013) for parasite resistance using the 50K-SNP array. Potential candidate genes in this region include: polycystin-2 (PKD2) and ATP binding cassette G member 2 (ABCG2), which have been reported as being under selection in a study of large number of breeds (Kijas et al., 2012). Although, ABCG2 has been investigated as a candidate gene for facial eczema in sheep (Duncan *et al.,* 2007).

Table 2: Summary of significant regions for RHM using 50 SNP window size analysis

OAR	Window start	Window finish	LRT	h_g^2	Candidate genes
2	105083320	107564404	10.07	0.0043	PALLD, DDX60, ANXA10
2	106585530	108470142	16.12	0.0045	
2	107564404	109633672	16.13	0.0048	
2	109633672	113113775	15.59	0.0051	
2	110827578	114955024	16.69	0.0051	
2	113113775	116350674	18.42	0.0084	
6	34614727	38019817	16.97	0.0083	PKD2, ABCG2, SP1
6	36522166	39035619	10.70	0.0054	
1	92157812	94722198	10.14	0.0048	CD58, CD2, CD101, IGSF3, VTCN1, FAM46C
3	129451837	1311779166	9.23	0.0047	SOCS2

CONCLUSION

This study has been successful at identifying QTLs for parasite resistance in a large multi-breed sheep population. The most significant regions were detected on chromosome 2 and 6. Four other regions on chromosome 1, 3 and 24 reached the suggestive threshold. These results also showed that there are a number of common genes that are underlying resistance to a wide range of parasite species. Furthermore, some of these common genes are possibly pleiotropic with other sheep diseases.

REFERENCES

Aulchenko Y. S., de Koning D-J., Haley C. (2007) *Genetics* **177**:577-585.
Butler, D.G., *et al.* (2007) ASReml-R reference manual. QLD Dept. of Primary Industries.
Dominik, S., Hunt, P.W., McNally, J et al. (2010) *Parasitology* **137**: 1275-1282.
Duncan, E., *et al.* (2007). *Animal genetics,* **38,** 126-131.
Hu Z-L., Park C. A., Wu X-L., Reecy J. M. (2013), *Nucleic Acids Res* **41**:D871–D879.
Jonas, E., *et al.* (2011) *Genetics Selection Evolution.* **43,** 22.
Kemper, K. E., Emery, D. L., Bishop, S. C. et al. (2011) *Genetics research* **93**:203-219.
Kijas, J. W., *et al.* (2012) *PLoS biology,* **10,** e1001258.
Marshall, K., Maddox, J.F., Lee, S.H., et al. (2009) *Animal Genetics*. **40**:262-272.
Nagamine, Y., Pong-Wong, R., Navarro, P. et al. (2012) *PLoS ONE* **7**:e46501.
Roeber, F., Jex, A. R., Gasser, R.B. (2013) *Parasites Vectors* **6**:153.
Van der Werf J.H.J.*et al.* (2010) *Anim.Prod.Sci.* **50:** 998.
Riggio, V., Matika, O., Pong-Wong, R. et al. (2013). *Heredity.* **110**:420-429.

Proc. Assoc. Advmt. Breed. Genet. 21: 73-76

NON-SYNONIMOUS POLYMORPHISM IN *HELB* IS ASSOCIATED WITH MALE AND FEMALE REPRODUCTIVE TRAITS IN CATTLE

M.R.S. Fortes[1], F.B. Almughlliq[1], L.T. Nguyen[1], L.R. Porto Neto[2] and S.A. Lehnert[2]

[1]University of Queensland, School of Chemistry and Molecular Biosciences, Qld 4072, Australia.
[2]CSIRO Agriculture Flagship, Queensland Bioscience Precinct, Qld 4067, Australia.

SUMMARY

We report on the association with reproductive phenotypes of a genetic marker mapped to the *HELB* gene, in tropically adapted beef cattle. The genetic marker is a single nucleotide polymorphism (SNP) in chromosome 5 that is non-synonymous and has a predicted deleterious effect on the coded protein. Reproductive phenotypes used in these analyses were from a population of the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC). Bulls (n = 1,023) were from a mixed-breed population and represented animals not included in the original genome-wide studies (GWAS) that reported SNP associations on chromosome 5 for scrotal circumference (SC). Cows (n = 1,089) were Tropical Composites and most were from the population used in the original GWAS that described SNP associations on chromosome 5 for puberty and post-partum anoestrus interval (PPAI). Animals were genotyped with Taqman assays designed for 3 non-synonymous SNP. The results indicate that the SNP in *HELB* is significant for SC and PPAI at $P < 10^{-7}$ and for heifer puberty ($P = 0.0029$). This SNP might aid across breed phenotype prediction for reproductive phenotypes given its significance in these mixed and composite cattle with both *Bos taurus* and *Bos indicus* ancestry. Future studies should target this *HELB* SNP in other breeds and populations to confirm the associations we described.

INTRODUCTION

Discovery of functional mutations related to QTL has merit in the context of animal breeding. Functional mutations, such as non-synonymous single nucleotide polymorphisms (SNP) or frameshift SNP have potential biological meaning: they might be the causative mutation or be in strong linkage disequilibrium (LD) with the causative mutation. For these reasons, functional mutations might overcome an important issue faced by genomic selection in the beef cattle industry: prediction of crossbred performance (Jonas and de Koning, 2015). Crossbred performance is especially important in tropical regions where beef cattle herds carry a mixture of *Bos indicus* and *Bos taurus* ancestry. Functional mutations associated with reproductive traits would especially benefit *Bos indicus* influenced cattle. *Bos indicus* breeds tend to have a prolonged post-partum anoestrus period, later puberty and overall lower reproductive performance when compared to *Bos taurus* breeds (Abeygunawardena and Dematawewa, 2004). Despite lower reproductive performance, *Bos indicus* breeds and crosses are largely used in tropical regions because of their adaptive advantage in those environments (Burrow, 2012). Improvement of reproductive performance of *Bos indicus* breeds and crosses would benefit beef production.

Herein, we investigated the association between 3 non-synonymous SNP in candidate genes mapped to *Bos taurus* autosome 5 (BTA 5) and reproductive phenotypes, in tropically adapted cattle. Previous studies have reported regions of BTA 5 to be associated with male and female reproductive phenotypes in various cattle breeds (Kappes et al., 2000; Kim et al., 2009; McClure et al., 2010; Fortes et al., 2012; Hawken et al., 2012; Fortes et al., 2013a). In Tropical Composite cattle, the QTL on BTA 5 was significant for scrotal circumference (SC) and post-partum anoestrus interval (PPAI). The SNP investigated in the present study are predicted deleterious mutations in *FAU*, *INHBC* and *HELB*, candidate genes mapped to this region of BTA 5. Our objective was to search for SNP that could explain the previously identified QTL.

MATERIALS AND METHODS

Animals. Blood for DNA extraction was obtained from 1021 bulls and 1089 cows. Bulls were 113 Brahman, 741 Tropical Composite and 167 crossbred, which were from the Beef CRC populations previously described (Burns et al., 2013). Importantly, the bulls used in this study had not been genotyped for previous GWAS (Fortes et al., 2012; Fortes et al., 2013a). Cows were Tropical Composites used in previous GWAS (Hawken et al., 2012). Crossbred bulls are the product of mating Brahman to Tropical Composites. Animals studied varied in *Bos indicus* and *Bos taurus* ancestry.

Mutations and genotype assays. Non-synonymous SNP were identified within three candidate genes - *FAU*, *INHBC* and *HELB* - by mining whole genome sequences of 64 bulls of a variety of breeds including Brahman and Senepol. Potentially deleterious mutations were identified using the bioinformatics software known as Variant Effect Predictor (VEP) from Ensembl (http://www.ensembl.org/info/docs/tools/vep/index.html). Custom TaqMan assays were developed for the selected non-synonymous SNP according to TaqMan Array Design Tool (Applied Biosystems, 2010). Further, bulls were also genotyped with the 90K Illumina SNP chip.

Analysis. Single SNP regression was applied for genotyped animals using a mixed model analysis of variance with the SNP & Variation Suite software (Release 8.3.0, Golden Helix, Inc.). The mixed model can be described with the equation: $y_i = X\beta + Z\mu + S_ja_j + e_i$; where yi represents the phenotypic measurement for the ith animal, X is the incidence matrix relating fixed effects (contemporary group and breed) in β with observations in y, Z is the incidence matrix relating to random additive polygenic effects of animal in μ with observations in y and Sj is the observed animal genotype for the jth SNP (coded as 0, 1 or 2 to represent the number of copies of the B allele), aj is the estimated SNP effect, and ei is the random residual effect. Age was fitted as a covariate for SC and PPAI. Same model for candidate SNP and GWAS.

RESULTS AND DISCUSSION

Previous QTL mapping and GWAS reported the importance of BTA 5 for cattle reproduction. A QTL for ovulation rate was mapped to 40 cM in the mixed-breed population known as MARC herd (Kappes et al., 2000). In Angus, a QTL for SC was reported at 13 Mb, at 104 Mb and at 127 cM (McClure et al., 2010). Associations with twinning rate for SNP between 55 and 75 Mb were reported in Holstein cattle (Kim et al., 2009). Hawken et.al (2012) carried GWAS in Brahman and Tropical Composites cows and found SNP associated with female reproductive traits in Tropical Composites located on BTA 5. The largest concentration of SNP associated with PPAI was on 44.0 to 44.3 Mb, 58.2 Mb and 113.6 Mb. Age of puberty was also reported to be associated with BTA 5 in 2 positions: 28.7 Mb and 96 Mb (Hawken et al., 2012). In Tropical Composite bulls, significant SNP association for levels of inhibin hormone were located between 42 and 61 Mb (Fortes et al., 2013a). Subsequently, significant SNP in BTA 5 were found for levels of inhibin and insulin-like growth hormone (IGF1) in Brahman and Tropical Composite bulls (Fortes et al., 2013b). Taken together the evidence points to one or more QTL on BTA 5 that could be important for reproductive physiology in many cattle breeds, with various *Bos indicus* and *Bos taurus* ancestry. In the current study, we tested 3 SNP in candidate genes located between 47 and 56 Mb of BTA 5, because these locations were important for tropically adapted cattle in the literature.

We have identified a SNP in *HELB* associated with SC and PPAI (Table 1, $P < 10^{-7}$). An association with heifer puberty, defined by the age at the first corpus luteum, was also noted ($P = 0.0029$). Breeding cows with the favourable allele could reduce PPAI in 30 days and puberty in 18 days. This SNP explained more than 2% of the genetic variance for SC and PPAI. The SNP in

FAU and *INHBC* failed to associate with studied phenotypes. Out of the 3 tested SNP, the one in *HELB* is the only mutation likely to be in high LD with the causative mutation in BTA 5. We considered the *HELB* SNP a potential functional mutation because it is non-synonymous and has predicted deleterious effect. It is a G to A substitution that causes a T to M residual change in the protein, with a SIFT score 41, deemed deleterious (0.04). The physiological effect of this protein change in SC, PPAI and AGECL is unknown and merits investigation. The causative mutation(s) in BTA 5 could be affecting the endocrine pathways that control both male and female reproduction; so that it is associated with SC, AGECL and PPAI. Previously, we observed an association between this QTL in BTA 5 and levels of IGF1 and inhibin (Fortes et al., 2013a). Links between this QTL and reproductive hormones should be explored.

Table 1. Estimated significance and effect of non-synonymous polymorphisms mapped to candidate genes on chromosome 5.*

	Phenotype	*Gene*	*P-Value*	*Effect*	*% Variance*
		FAU	$2.82x10^{-1}$	0.3507	0.12
	SC18	***HELB***	$\mathbf{5.23x10^{-7}}$	**0.7018**	**2.50**
Bulls		*INHBC*	$7.86\ x10^{-1}$	-0.0368	0.01
		FAU	$5.77\ x10^{-1}$	0.1724	0.03
	SC24	***HELB***	$\mathbf{1.94x10^{-11}}$	**0.8949**	**4.40**
		INHBC	$4.22x10^{-1}$	-0.1044	0.06
		FAU	$7.67x10^{-1}$	3.9005	0.01
	AGECL	***HELB***	$\mathbf{2.90x10^{-3}}$	**18.3899**	**0.95**
		INHBC	$8.47x10^{-1}$	1.1082	0.00
Cows		*FAU*	$7.68x10^{-1}$	3.4762	0.01
	PPAI	***HELB***	$\mathbf{2.15x10^{-8}}$	**30.4043**	**3.87**
		INHBC	$8.53x10^{-1}$	-0.9703	0.00

*Genes and values highlighted in bold represent significant associations for reproductive traits: scrotal circumference at 18 and 24 months of age (SC18 and SC24, cm), age at first corpus luteum (AGECL, days) and post-partum anoestrus interval (PPAI, days). The SNP ***Effect*** is provided in the same measuring unit at the trait and ***% Variance*** is the percentage of the additive genetic variance explained by each SNP.

The relevance of the SNP in *HELB* to previously described QTL was tested by fitting this SNP as a fixed effect in the GWAS (Figure 1). When comparing the *P-values* between two models, we observe that the *P-values* are reduced from 10^{-8} to 10^{-6} for SC18, and from 10^{-13} to 10^{-10} for SC24. However, the strong association signal from common chip SNP was still present after fitting the *HELB* SNP. This result is different from expectations for causative mutations. When the causative mutation in the *PLAG1* gene was fitted the QTL on BTA 14 practically disappeared from GWAS (Fortes et al., 2013c). This result suggests that more in-depth molecular characterisation is required. It is unlikely that the SNP in *HELB* is the only causative mutation associated to reproductive traits in this genomic region.

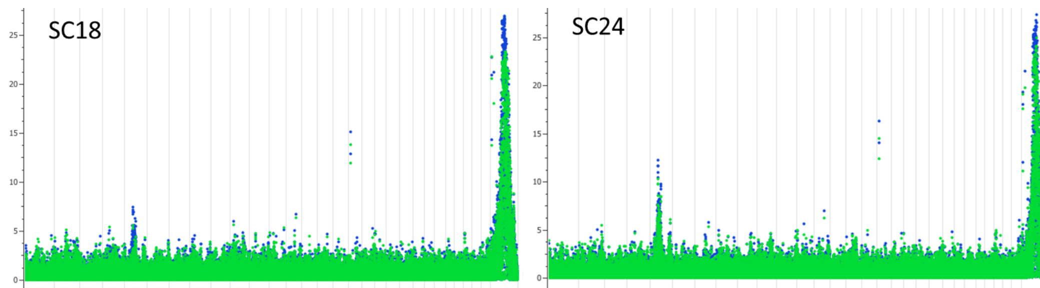


Figure 1. Significance of the associations with SC18 and SC24 for SNP across the genome. *P-values* are on y-axis and genomic positions are on x-axis. Note the reduction in *P*-values from the blue dots (animal model) to the green dots (animal model with *HELB* SNP fitted as a fixed effect).

CONCLUSION

The identification of a SNP associated with SC in a mixed population of bulls and with PPAI and AGECL in Tropical Composite cows might aid across breed phenotype predictions. Future studies should target this *HELB* SNP in other breeds and populations to confirm associations. For female traits, this SNP needs to be validated independently from the original GWAS.

ACKNOWLEDGEMENTS

The authors acknowledge all collaborators of the Beef CRC. Meat and Livestock Australia provided support (project B.NBP.0786). Fortes is supported by UQ Postdoctoral Fellowship.

REFERENCES

Abeygunawardena H. and Dematawewa C.M. (2004) *Anim. Reprod. Sci.* **83**: 373-387.

Burns B.M., Corbet N.J., Corbet D.H., Crisp J.M., Venus B.K., Johnston D.J., Li Y., McGowan M.R. and Holroyd R.G. (2013) *Anim. Prod. Sc.* **53**: 87-100.

Burrow H.M. (2012) *Anim.* **6**: 729-740.

Fortes M.R.S., Reverter A., Hawken R.J., Bolormaa S. and Lehnert S.A. (2012) *Bio. Reprod.* **87**: 1-8.

Fortes M.R.S., Reverter A., Kelly M., McCulloch R. and Lehnert S.A. (2013a) *Androl.* 1: 644-650.

Fortes M.R.S. Reverter A., Porto Neto L.R., Kelly M., Moore S.S. and Lehnert S.A. (2013b) *Proc. Ass. Adv. Anim. Breed. Genet.* **20**: 389-392.

Fortes M.R.S., Kemper K., Sasazaki S., Reverter A., Pryce J.E., Barendse W., Bunch R., McCulloch R., Harrison B., Bolormaa S., Zhang Y.D., Hawken R.J., Goddard M.E. and Lehnert S.A. (2013c) *Anim. Genet.* **44**: 636-647.

Hawken R.J., Zhang Y.D., Fortes M.R.S., Collis E., Barris W.C., Corbet N.J., Williams P.J., Fordyce G., Holroyd R.G., Walkley J.R.W., Barendse W., Johnston D.J., Prayaga K.C., Tier B., Reverter A. and Lehnert S.A. (2012) *J. Anim. Sci.* **90**: 1398-1410.

Jonas E., and D.J. de Koning. (2015) *Front. Genet.* **6**: 49.

Kappes S.M., Bennett G.L., Keele J.W., Echternkamp S.E., Gregory K.E. and Thallman R.M. (2000) *J. Anim. Sci.* **78**: 3053-3059.

Kim E.S., Berger P.J. and Kirkpatrick B.W. (2009) *J. Anim. Sci.* **87**: 835-843.

McClure M.C., Morsci N.S., Schnabel R.D., Kim J.W., Yao P., Rolf M.M., McKay S.D., Gregg SJ, Chapple R.H., Northcutt S.L. and Taylor J.F. (2010) *Anim. Genet.* **41**: 597-607.

Proc. Assoc. Advmt. Breed. Genet. 21: 77-80

DETECTING VARIANTS ASSOCIATED WITH COMPLEX TRAITS THROUGH CHANGING GENE EXPRESSION IN CATTLE

M. Khansefid[1,2,3], J.E. Pryce[2,3,4], S. Bolormaa[2], Y. Chen[5] and M.E. Goddard[1,2,3]

[1] Department of Agriculture and Food Systems, The University of Melbourne, VIC, Australia
[2] Department of Economic Development, Jobs, Transport and Resources, VIC, Australia
[3] Dairy Futures Cooperative Research Centre (CRC), VIC, Australia
[4] La Trobe University, VIC, Australia
[5] Department of Primary Industries, Elizabeth Macarthur Agricultural Institute, NSW, Australia

SUMMARY

Mutations can affect phenotypes by changing the amino acid sequence in a protein or by changing the expression of a gene. The gene expression in a particular tissue can be measured using mRNA sequencing and counting the number of mRNA copies of the gene. The aim of our research was to find mutations which affect expression of genes in cis. We detected the mutations that are associated with the expression of each gene in muscle (45 Angus bulls) and liver (38 Angus bulls) by correlating the mRNA count with the alleles carried at single nucleotide polymorphisms (SNPs) within 50kb of the genes that were tested. Furthermore, the SNPs and genes with at least one SNP significantly associated with one or more traits ($p < 0.001$) were found by genome-wide association studies (GWAS) in a beef cattle dataset including 6,114 genotyped animals with 20 traits recorded. We compared the results to find the SNPs significantly associated with gene expression ($p < 0.001$) and also with the variation in phenotype. The SNPs which were significantly associated with gene expression ($p < 0.001$) were more likely to be significant in GWAS for concentration of Insulin like growth factor1(IGF-1), residual feed intake (RFI) and in a multi-trait significance test.

INTRODUCTION

The mutations underlying variation in complex traits are generally identified using genotypes and phenotypes in GWAS (Goddard and Hayes 2009). However, most of the SNPs found in GWAS are not located in the coding regions of the genome and so may influence the variation of the traits by changing the expression of genes. These polymorphic sites are likely to affect the expression of the copy of the gene on the same chromosome and are called cis-expression quantitative trait loci (cis-eQTL) (Arnold *et al.* 2012).

The next generation sequencing (NGS) technology which is used for sequencing DNA is also applicable for sequencing mRNA. In genome-wide transcriptome studies using the mRNA sequencing (RNA-seq) data, the abundance of mRNA for each gene can be measured by counting the number of sequence reads aligned to the gene in the reference genome.

Although GWAS have found many associations between SNPs and traits, it has proved difficult to identify the causal mutation or the gene(s) whose structure or expression they affect. On the other hand, a cis-eQTL affects the expression of a specified gene, so if a cis-eQTL is the same as a QTL for a traditional phenotype, this defines the gene through which the QTL probably acts. In addition, the cis-eQTL may help to identify the causal mutation for both itself and the traditional QTL. However, the RNA-seq data is more expensive to obtain than most traditional phenotypes and therefore the number of animals in whole transcriptome studies is limited. Also, the results depend on the tissue which was sampled for mRNA and so, even if a QTL is identical with a cis-eQTL, this may not be discovered if the wrong tissue is sampled.

The aim of our research was to find if SNPs which were significantly associated with gene expression (cis-eQTL) were more likely to be significant in GWAS of traditional traits.

MATERIALS AND METHODS

Animals. The animals used for RNA-seq were young Angus bulls from lines of cattle divergently selected for residual feed intake (RFI) from the selection lines established in 1993 at the Agricultural Research Centre, Trangie, NSW, Australia (Arthur *et al.* 2001). The bulls with liver (L-bulls) and muscle (M-bulls) samples were from the lines after approximately 3 and 4 generations of selection, respectively.

RNA-seq. The transcriptome data was for 43 muscle and 38 liver samples. The extracted RNA from the sampled tissues (Chen *et al.* 2011) were sequenced with a HiSeq 2000 (Illumina Inc) after mRNA enrichment, by the modified protocol of Illumina sample preparation for RNA-seq. All of the raw reads were passed through quality control filters and trimming the ends of reads based on Phred quality scores (minimum quality of each nucleotide base=15, minimum average quality of the read after trimming=20, minimum read length after trimming=50 and maximum consecutive nucleotide bases with poor quality=3).

Genotypes. For 43 M-bulls we had 800K SNP chip and whole genome sequence data (WGS) with average coverage 6.7 fold from the 1000 bull genomes project (Daetwyler *et al.* 2014). The L-bulls had 800K SNP genotypes imputed from 50K (Illumina BovineSNP50K chip) using BEAGLE (Browning and Browning 2009) and then imputed to WGS using FImpute (Sargolzaei *et al.* 2014). We also used FImpute to rephase the genotypes of M and L bulls.

Alignment. The bovine genome assembly UMD3.1 was modified using the WGS data to produce a genome sequence for each individual bull. For each bull the RNA-seq data were aligned to its customised reference genome using TopHat2 (Kim *et al.* 2013).

Abundance of genes. For each gene in the reference genome, the mapped mRNA was counted using HTSeq python package (Anders *et al.* 2015). The number of sequence reads for all genes in each animal were normalised with a weighted trimmed mean of the log expression ratios using edgeR package in R (Robinson and Oshlack 2010). Finally, the normalised gene counts were log transformed to have normal distributions across animals.

eQTL mapping. The genes expressed in more than 25% of L-bulls and M-bulls were used to find eQTL. The association between the gene counts and the SNPs in WGS data within 50kb of the gene was calculated with ASReml for muscle and liver samples separately.

GWAS. A GWAS was carried out for 20 traits (including meat quality and production traits) using up to 6114 cattle and 729,068 HD SNPs. As well as individual traits, a multi-trait test was performed as described by Bolormaa et al (2014), except that only *Bos taurus* cattle were included. The SNPs were from the Illumina high density panel and were either genotyped or imputed from lower density. Only SNPs within 50 kb of a gene were used so that the same SNPs were tested in the GWAS as were tested for cis-eQTL.

The SNPs were classified as significant or not for association with gene expression ($p < 0.001$) and for association with one of the traits ($p < 0.001$) and we performed a chi-squared test of the hypothesis that SNPs affecting gene expression are more likely to be significant in GWAS. We also classified genes as either containing a significant cis-eQTL or not and as having a SNP within 50kb associated with a trait or not, and performed a chi-square test to test the hypothesis that genes containing an eQTL were also likely to be near a SNP associated with a traditional phenotype.

RESULTS AND DISCUSSION

RNAseq. In the muscle samples, there were on average about 8.5×10^6 (100%) RNA-seq raw reads per animal, 6.5×10^6 (75%) reads that passed the quality control filters, 5.9×10^6 (70%) that aligned to the reference genome and 5.5×10^6 (65%) that were mapped uniquely. In the liver samples, there were on average roughly 7.6×10^6 (100%) raw reads per animal, 5.5×10^6 (72%) reads that passed the quality control filters, 4.6×10^6 (60%) that aligned to the genome and 4.5×10^6 (52%) that were mapped uniquely. The percentage of reads mapped to the reference genome in L-

bulls is less than M-bulls probably because we used imputed genotypes to enhance the L-bulls reference genomes. In all chromosomes (and autosomes), 12,278 (11,842) in muscle and 12,233 (11,821) genes in liver were expressed (about 50% of the known genes were expressed in muscle and liver).

cis-eQTL.. In the muscle samples, each of the expressed 12,278 genes were tested for cis-eQTL using SNPs within 50kb of the gene. Among the HD SNPs, there were 240,818 SNPs that were tested for association with expression of one or more genes. 5,042 of these SNPs were significantly ($p < 0.001$) associated with expression of at least one gene in muscle. Similarly, of 227,488 SNPs tested, 2,420 were associated ($p < 0.001$) with expression of at least one gene in liver (Table 1).

Trait QTL. Table 1 presents results for two individual traits (blood concentration of IGF1 and residual feed intake) and the multi-trait test. For instance, 1,047 SNPs, out of 240,818 tested, were significantly associated with RFI (The number of SNPs tested varies slightly between M and L bulls and by trait because some SNPs had to be dropped from some analyses because they had a MAF below 1% in that dataset and only SNPs near genes expressed in that tissue were used).

Overlap between trait QTL and eQTL. There were only 3 SNP in common between the 386 SNPs that were associated with blood concentration of IGF-1 and the 5,041 associated with expression of at least one gene in muscle. This is not more than expected by chance (p=0.07) (Table 1). However, as shown in Table 1, there was more overlap between SNPs associated with traits and gene expression than expected by chance ($p < 0.05$) in 3 of the 6 tests. For instance, there were 169 SNPs associated with both the multi-trait test and with gene expression in muscle and this was far more than expected by chance ($p=8.2\times10^{-9}$).

Where a SNP is significantly associated with a trait and with expression of a gene, the expression of this gene may also be affecting the trait. Genes identified in this way (and the number of significant SNPs associated with their expression) for the cis-eQTL in muscle and affecting IGF1 are: *PPM1H* (2), *MTHFD1* (1). For muscle cis-eQTL and RFI: *POLR2I* (33), *PTPRR* (5), *ATP5E* (4), *SSFA2* (2), *CARD6* (2), *THAP8* (1), *DNER* (1), *ATPIF1* (1) and in cis-eQTL in liver and IGF-1: *DAB1* (2), *EVC* (1), RFI: *MVK* (9), *GSK3A* (4), *LOXL3* (1), *DPYD* (1), *DNER* (1), and 2 SNPs were in an uncharacterized gene. For example, *GSK3A* (*glycogen synthase kinase 3 alpha*) was reported to regulate glycogen metabolism in liver (Ali *et al.* 2001). So it seems reasonable that this gene can affect feed efficiency and RFI. The number of SNPs found significant in eQTL mapping and in the multi-trait GWAS was 169 in muscle and 48 in liver and these SNPs were spread across 51 genes in muscle and 25 in liver.

Table 1 also contains the corresponding numbers based on testing genes instead of SNPs. 4,044 genes had a cis-eQTL, 303 genes had a SNP within 50kb associated with IGF-1 and 125 genes had both which was significantly more than expected by chance (p=0.002). Note that the SNPs associated with IGF-1 are not necessarily the same as the SNPs associated with gene expression, they are just near the same gene. In addition, the eQTL analysis used whole genome sequence whereas the QTL analysis used HD SNPs.

In all 6 tests, there were more genes that contained both a QTL and an eQTL than expected by chance (Table 1). Examples of genes that contain a QTL and an eQTL are: in muscle, *SH3-domain GRB2-like (endophilin)-interacting protein 1* (*SGIP1*) gene was previously reported to have a role in regulating food intake, fat mass, energy balance and energy homeostasis. Its roles in regulation of feeding behaviour affects any process that activates or increases the frequency, rate or extent of feeding behaviour (Trevaskis *et al.* 2005; Cummings *et al.* 2012) and therefore might affect RFI. *Bos taurus bone morphogenetic protein 2 (BMP2)* is one of the genes found to significantly associated with IGF-1 concentration and differentially expressed in muscle. BMPs, are also called growth and differentiation factors have negative regulation of the IGF receptor signalling pathway and affect any process that stops, prevents, or reduces the frequency, rate or extent of IGF receptor

signalling (Kronenberg 2003).

We conclude that traditional QTL are sometimes in fact cis-eQTL. Therefore mapping cis-eQTL will help us to identify causal variants for conventional phenotypes and the genes through which these variants act. A benefit of eQTL is that the gene whose expression they affect is known so, if the QTL is an eQTL, this identifies the gene through which the QTL acts. cis-eQTL often explain a large proportion of the variance in expression and so there is some power to identify the causal variant even in small datasets. In addition, as information builds up about regulatory regions in livestock genomes, we will have functional information to help us identify sites that might change the expression of the target gene.

Table 1. SNPs and genes association with traits variation and gene expression in muscle and liver

Tissue / Trait	Total SNPs (genes)	Number of SNPs (genes) significantly associated with:			
		QTL	eQTL	QTL & eQTL	X^2 p-value for SNPs (genes)
Muscle					
IGF-1	240,586 (12,278)	386 (303)	5,041 (4,044)	3 (125)	0.070 (0.002)
RFI	240,818 (12,278)	1,047 (502)	5,042 (4,044)	49 (204)	4.6×10^{-09} (1.8×10^{-04})
Multi-Trait	239,726 (11,842)	5,240 (3,102)	5,030 (3,923)	169 (1,099)	8.2×10^{-09} (0.002)
Liver					
IGF-1	227,473 (12,233)	366 (287)	2,420 (2,246)	3 (85)	0.649 (6.2×10^{-07})
RFI	227,488 (12,233)	985 (497)	2,420 (2,246)	18 (113)	0.019 (0.010)
Multi-Trait	226,564 (11,821)	4,907 (2,973)	2,413 (2,202)	48 (630)	0.550 (3.3×10^{-05})

REFERENCES

Ali A., Hoeflich K.P., and Woodgett J.R. (2001) *Chem. Rev.* **101**: 2527.

Anders S., Pyl P.T. and Huber W. (2015) *Bioinformatics* **31**:166.

Arnold M., Ellwanger D.C., Hartsperger M.L., Pfeufer A. and Stümpflen V. (2012) *PLoS One* **7**:e36694.

Arthur P.F., Archer J.A., Johnston D.J., Herd R.M., Richardson E.C., *et al.* (2001) *J. Anim. Sci.* **79**:2805.

Bolormaa S., Pryce J. E., Reverter A., Zhang Y., Barendse W., *et al.* (2014). *PLoS Genet.* **10**: e1004198.

Browning B.L. and Browning S.R. (2009) *Am. J. Hum. Genet.* **84**:210.

Chen Y., Gondro C., Quinn K., Herd R.M., Parnell P.F., *et al.* (2011) *Anim Genet* **42**:475.

Cummings N., Shields K.A., Curran J.E., Bozaoglu K., Trevaskis J., *et al.* (2012) *Int. J. Obes.* **36**:201.

Daetwyler H. D., Capitan A., Pausch H., Stothard P., Van Binsbergen R., *et al.* (2014) *Nature Genet.* **46**:858.

Goddard M.E. and Hayes, B.J. (2009) *Nature Rev. Genet.* **10**: 381.

Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., *et al.* (2013) *Genome Biol* **14**:R36.

Kronenberg H.M. (2003) *Nature* **423**:332

Robinson M.D. and Oshlack A. (2010) *Genome Biol* **11**:R25.

Sargolzaei M., Chesnais J.P. and Schenkel F.S. (2014) *BMC Genomics* **15**:478.

Trevaskis J., Walder K., Foletta V., Kerr-Bayles L., McMillan J., *et al.* (2005) *Endocrinology* **146**:3757.

Proc. Assoc. Advmt. Breed. Genet. 21: 81-84

MULTI-BREED GWAS AND META-ANALYSIS USING WHOLE-GENOME SEQUENCES OF FIVE DAIRY CATTLE BREEDS

I. van den Berg[1,2,3], D. Boichard[2] and M.S. Lund[1]

[1] Center for Quantitative Genetics and genomics, Department of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark
[2] INRA, UMR1313 GABI, 78350, Jouy-en-Josas, France
[3] AgroParisTech, Paris, France

SUMMARY

A multi-breed genome wide association study (GWAS) can potentially improve QTL mapping precision and detection power. Alternatively to a multi-breed GWAS, meta-analysis can combine within breed GWAS results. Our objective was to compare within breed GWAS, multi-breed GWAS and meta-analysis of within breed GWAS results. Imputed whole-genome sequences and deregressed proofs for milk, fat and protein yield of 16,031 bulls of five French and Danish dairy cattle breeds were used for the analyses. GWAS were performed within each breed, combining French and Danish Holstein, combining Jersey, Montbéliarde, Normande and Danish Red, and combining all breeds. Within breed GWAS results were combined using three different meta-analysis models. The multi-breed GWAS resulted in more distinct peaks by increasing the p-values of some variants and decreasing the p-values of others. For some QTL not segregating in Holstein, combining all breeds except Holstein was useful, because they were overshadowed by larger QTL segregating in Holstein when all breeds were combined. The meta-analysis gave results similar to the multi-breed GWAS and can be used as an alternative. The results obtained by the weighted Z-score model were closest to those of the multi-breed GWAS.

INTRODUCTION

Genome wide association studies (GWAS) can help in the identification of causative mutations influencing quantitative traits. With the increasing number of re-sequenced individuals, more causative mutations are directly present in the data. In addition, however, there is also a large number of variants in linkage disequilibrium (LD) with the causative mutations. As a consequence, especially in populations with high levels of long range LD, as is the case within dairy cattle breeds (de Roos *et al*., 2008), GWAS generally results in large number of variants associated with a QTL, over a large region. Across breed, LD is only shared for short distances, and multi-breed GWAS could therefore improve QTL mapping precision. Furthermore, with the large number of sequence variants, high thresholds are necessary to avoid too many false positives. For breeds with small study populations, the detection power of a within breed GWAS might not be sufficient to detect QTL with a small effect. If causative mutations are shared across breed, a multi-breed GWAS could help to improve detection power and aid the identification of such QTL.

A multi-breed GWAS could thus potentially improve both mapping precision and detection power. It is, however, not always possible to have all data required for a multi-breed GWAS. Alternatively, a meta-analysis can be performed, that combines results of individual GWAS (Begum *et al*., 2012). In human, Lin and Zeng (2010) found similar efficiency for a meta-analysis as for a full joint analysis.

Our objective was to compare different multi-breed GWAS approaches, using whole-genome sequence data of five French and Danish dairy cattle breeds. GWAS was performed both within breed and multi-breed, and three meta-analysis methods were compared to the multi-breed GWAS.

MATERIALS AND METHODS

Imputed sequences of 4993 Danish Holstein, 984 Jersey, 768 Danish Red, 5626 French Holstein, 1935 Montbéliarde and 1725 Normande bulls and deregressed proofs obtained following Garrick *et al.* (2009) for milk yield, fat content and protein content were used for the analyses. First, bulls genotyped with the 50K chip were imputed to HD. For the French data (Hozé *et al.*, 2013), this step was performed using Beagle 3.0.0 (Browning and Browning, 2007), while for the Danish breeds, IMPUTE2 was used (Howie *et al.*, 2009). Subsequent imputation to whole-genome sequence was for all breeds done using IMPUTE2. The reference used for imputation to sequences of the Danish bulls consisted of the bulls in run 4 of the 1000 bull genome project (Daetwyler *et al.*, 2014), while for the imputation of the French bulls, a combined French-Danish reference set was used. The latter consisted of 122 Holstein, 27 Jersey, 28 Montbéliarde, 23 Normande and 45 Danish Red bulls. In total, 24,550,115 polymorphisms were used for the analysis, after filtering for imputation quality (IMPUTE2 info score ≥ 0.6) and minor allele frequency (MAF) (≥ 0.005).

To study genomic relationships between breeds, a genomic relationship was constructed using SNP from the 50K chip for 500 randomly selected individuals of each breed. Genomic relationships were standardised and scaled based on allele frequencies estimated in the animals used to construct the genomic relationship matrix, following VanRaden (2009). Subsequently, a principal component analysis (PCA) was performed using the prcomp() command in R (2015).

A GWAS was performed within each breed, using a single marker model with a random sire effect:

$$y_{ij} = \mu + S_j + \beta g_{ij} + e_{ij} ,$$

where y_{ij} is the DRP for individual i with sire j, S the random effect of sire j, b the effect of the polymorphisms, g_{ij} the allele dose (ranging from 0 to 2) of individual i with sire j and e_{ij} a random residual.

Afterwards, for all variants with a within breed p-value below 10^{-5} in French or Danish Holstein or below 10^{-3} in one of the other breeds for at least one trait were used for the multi-breed GWAS. The multi-breed GWAS was performed combining French and Danish Holstein (HOL), combining Jersey, Danish Red, Montbéliarde and Normande (REST), and combining all populations (ALL). The model used was identical to that used within breed, except for the addition of a breed effect.

Three meta-analysis approaches were used to combine within breed GWAS results: the weighted Z-scores model using METAL software (Willer *et al.*, 2010), and the fixed and random effects models using META software (Liu *et al.*, 2010). The inputs of the Z-score model are within breed p-values, effect direction and sample size, while the fixed and random effects models use the within breed effects and standard errors. The random effects model accounts for heterogeneity between studies using Cochran's statistic.

RESULTS AND DISCUSSION

Figure 1 shows the genomic relationship between the different breeds used for the studies. French and Danish Holstein populations were very similar, and Danish Red was closer than Montbéliarde and Normande, while Jersey was the most distinct from the other breeds.

The multi-breed GWAS generally resulted in more distinct peaks than the individual within breed GWAS. When only the two Holstein populations were combined, p-values decreased due to the larger detection power. When all breeds or all breeds except Holstein were combined, p-values of some variants decreased, but increased for others.

For QTL segregating in multiple breeds, adding more breeds resulted in stronger associations and decreased p-values. Peaks became more distinct when more different breeds were added, also for QTL that were segregating in only one or few breeds. For such QTL, the p-values of variants segregating in the breeds where the QTL is not present increased. When, however, a region

contained different QTL segregating in different breeds, QTL segregating in breeds with a smaller sample size were sometimes overshadowed by QTL segregating in Holstein.

Figure 2 shows a peak around 94 Mb on chromosome 5 associated with fat yield in Holstein. Within breed, the peak was present in both Holstein populations with a $-\log_{10}(p)$ around 33, and a smaller peak on the same location was detected in Normande. Combining the two Holstein populations increased the $-\log_{10}(p)$ of the top variant to 62.6, and adding the other breeds resulted in a further increase in the peak. The most significant variant had a $-\log_{10}(p)$ of 71.6, and was an intron in MGST1, with rs-id rs211210569, a gene known ~~of~~ for its association with fat yield (Raven *et al*., 2014).

In the other breeds, several peaks were detected in the same region. In the multi-breed GWAS combining all breeds, these peaks seem to disappear due to the large peak in MGST1. When all breeds except Holstein were combined, however, a clear peak was detected around 112.5 Mb, as shown in figure 4. Within breed, this peak was observed in Normande and Jersey. The most significant variant in the multi-breed analysis excluding Holstein was an intron in MKL1, with rs-id rs110294643. MKL1 plays an important role in mammary gland development in mice (Sun *et al.*, 2006).

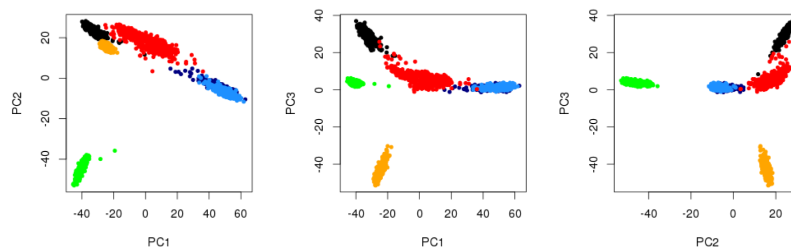


Figure 1. Principal component analysis of genomic relationships. Showing principal components (PC) 1, 2 and 3, dark blue = Danish Holstein, light blue = French Holstein, green = Jersey, black = Montbéliarde, orange = Normande, red = Danish Red.

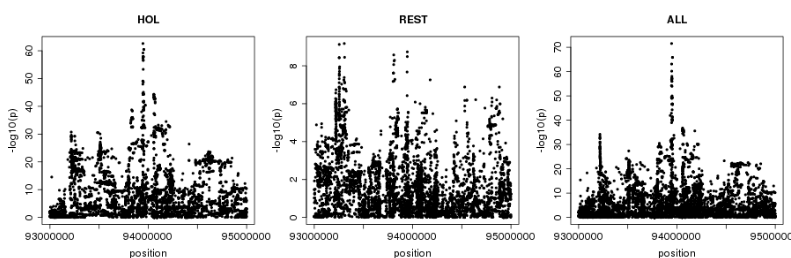


Figure 2. $-\log_{10}(p)$ for fat yield in the multi-breed analysis on chromosome 5 (93-95Mb)

Table 1 gives the correlation between p-values obtained in the multi-breed analysis and those obtained in the different meta-analyses. The weighted Z-score model gave the most similar results to the multi-breed GWAS. The weighted Z-scores model uses p-values as input rather than estimated effects, and is therefore less influenced by scaling differences. The random effects model gave for some variants very similar results to the multi-breed GWAS. For a large part of the variants, however, heterogeneity detected by this model was large, resulting in high p-values, even for variants that showed strong associations in the multi-breed analysis. All meta-analyses gave more different results from the multi-breed GWAS when different breeds were combined than

when the two Holstein populations were combined. Not all QTL are segregating in all breeds, and as a consequence, it is more difficult to estimate an overall effect in a multi breed analysis.

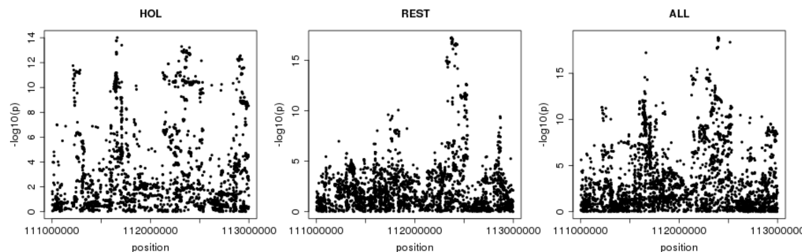


Figure 3. $-\log_{10}(p)$ for fat yield in the multi-breed analysis on chromosome 5 (111-113Mb)

Table 1. Correlations between p-values obtained in multi-breed analysis and p-values obtained by meta-analysis for variants with a p-value below 10^{-5} in Holstein or 10^{-3} in Jersey, Montbéliarde, Normande or Danish Red in a within breed GWAS

	milk			fat			protein		
	Z	**F**	**R**	**Z**	**F**	**R**	**Z**	**F**	**R**
HOL	0.97	0.87	0.84	0.97	0.88	0.86	0.96	0.87	0.85
REST	0.54	0.26	0.45	0.81	0.70	0.78	0.90	0.79	0.85
ALL	0.48	0.28	0.34	0.85	0.69	0.57	0.86	0.70	0.57

Z = weighted Z-scores, F = fixed effects and R= random effects

CONCLUSIONS

The multi-breed analysis helped to improve the precision of QTL mapping compared to the within breed GWAS. However, due to the much larger number of records available for Holstein than for the other breeds, when different QTL are segregating in different breeds in the same region, the Holstein QTL tended to dominate the results. Combining all breeds except Holstein was therefore useful to detect some QTL segregating in the other breeds that were overshadowed by larger Holstein QTL. A meta-analysis can be used as an alternative for a full multi-breed analysis. The weighted Z-score model gave results most similar to those of the multi-breed GWAS.

ACKNOWLEDGEMENTS

IB benefited from an Erasmus-Mundus fellowship and a grant by Apisgene, within the framework of the European Graduate School in Animal Breeding and Genetics.

REFERENCES

Begum F., Gosh D., Tseng G.C. and Feingold E. (2012) *Nucleic Acids Res.* **40**:3777.

de Roos A.P.W., Hayes B.J., Spelman R.J., and Goddard M.E. (2008) *Genetics* **179**:1503.

Garrick D.J., Taylor J.F. and Fernando R.L. (2009) *Genet. Sel. Evol.* **41**:55.

Lin D.Y. and Zeng D. (2010) *Genet. Epidemiol.* **34**:60.

Liu J.Z., Tozzi F., Waterworth D.M., Pillai S.G., Muglia P. *et al.* (2010) *Nat Genet* **42**:436.

R Core Team (2015) http://www.R-project.org/

Raven, L-A., Cocks B.G. and Hayes B.J. (2014) *BMC Genomics* **15**:62.

Sun Y., Boyd K., Xu W., Ma J., Jackson C.W. *et al.* (2006) *Mol. Cell. Biol.* **26**:5809.

Willer C.J., Li Y. and Abecasis C.R. (2010) *Bioinformatics* **26**:2190.

Proc. Assoc. Advmt. Breed. Genet. 21: 85-88

VISUALISATION OF RUMEN MICROBIOME DATA

G.R. Wood[1], S. Kittelmann[2], C. Pinares-Patiño[2], K.G.Dodds[1], M.R. Kirk[2], S. Ganesh[2], S.M. Hickey[3], P.H. Janssen[2], J.C. McEwan[1] and S.J. Rowe[1]

[1]AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand
[2]AgResearch, Grasslands Research Centre, Palmerston North, New Zealand
[3]AgResearch, Ruakura Research Centre, Hamilton, New Zealand

SUMMARY

The rumen microbiome plays a key role in the production of methane, a critical greenhouse gas. Two graphical methods aiding visualization of rumen microbiome data are presented. Marker gene sequence data generated from 520 rumen samples from 260 sheep were used to estimate the relative abundance of bacteria. These were assigned to 54 bacterial taxa, which were then clustered into four groups of co-occurring taxa. Sheep were measured for methane emissions in open circuit respiration chambers. Heatmaps and simplex plots were used to visualise the rumen data and their relationship to methane emissions from the sheep. Although the relationship of these dimensions to methane emission was not clear-cut, the simplex graphic suggests that there is a continuum of low methane emitters across rumen profiles. The analysis indicated that the bacterial microbiome data set is broadly two-dimensional; two rumenotypes dominate. This in turn creates a challenge - to uncover the origins of this relative simplicity of rumen biology.

INTRODUCTION

Methane emission from ruminants (primarily sheep and cattle) contributes approximately 30% of annual anthropogenic greenhouse gas (GHG) production in New Zealand. Methane is produced in a two-stage process in the rumen: feed digestion aided by rumen bacteria yields hydrogen as a by-product, which in turn is converted to methane by methanogenic archaea.

The bacterial composition of the rumen microbiome can be determined through sequencing part of the gene coding for a slowly evolving ribosomal RNA. Sequenced sections are aligned to reference sequences of the gene in different bacteria and then assigned to bacterial taxa. This approach resulted in 54 taxa in the data set used in this study. The relative abundance of these taxa in the rumen of each animal can then be estimated. Taxa, once established, can be clustered if their relative abundances vary in concert across animals; in turn, this allows clustering of taxa. These co-occurring taxa are not necessarily related evolutionarily – instead, they are probably linked ecologically. A final research aim, of which the current research is a part, is to determine an animal measure which can be used to select for lower methane emission. It is known that animal genetics influences methane production (Pinares-Patiño *et al*. 2013) and that the rumen microbiome influences methane production (Kittelmann *et al*. 2014; Ross *et al*. 2013; Wallace *et al*. 2014).

The microbiome consists of multiple microbial species, each representing a dimension in the data set. Microbiome compositions within large animal groups are unlikely to form discrete clusters (where each point is a rumen sample), and so methods for simplifying these large data sets are useful in order to make progress. One route is to reduce the dimensionality – from 54 to a workable number. Classical methods are available, such as correspondence analysis. The aim here, however, is to present two simple visual methods, first a marriage of cluster analysis and heatmaps and second (related to correspondence analysis, (Greenacre 1983)) a marriage of cluster analysis and simplex plots. Each presents graphically, but in different ways, the relationship between animals, microbiome and methane production.

It is stressed that this paper focuses on exploratory graphical methods, designed to allow the

researcher to "see", and hence conjecture, relationships; confirmatory standard statistical tests are not conducted. Graphical methods are readily accessible; two examples are presented here with a view to providing useful tools to aid the progress of microbiome research.

MATERIALS AND METHODS

A full description of the selection lines used together with the materials and methods for measuring methane emissions can be found in Pinares-Patiño *et al*. (2011). Selection began in 2009 and the sample of sheep used in this study were from the 2011 and 2012 birth cohorts. Two rumen contents samples were collected by stomach tube from each of 260 New Zealand crossbred sheep at ~10 months of age. Rumen samples were collected 14 days apart approximately 18 hours after the last feed. The sheep were from selection lines phenotypically divergent for methane emissions per unit of dry matter intake (CH_4/DMI; g/kg) and were measured in open circuit respiration chambers for methane emissions continuously for 48 hours prior to each rumen sampling Pinares-Patiño *et al.* (2013). Sheep were offered a ration of lucerne pellets based on 2.0× maintenance energy requirements. Feed was offered twice daily and individual dry-matter intakes were recorded. Breeding values were estimated using ASReml 3 software (Gilmour 2009).

The method for assigning amplified 16S rRNA gene sequences to taxa is described by Kittelmann *et al*. (2014). The statistical package R was used for clustering and heatmaps and Matlab for the production of rotatable three-dimensional simplex plots (two representative screenshots are presented here). Simplexes can be drawn in each dimension; one-dimensional (line), two-dimensional (triangle); three-dimensonal (tetrahedron), etc. They provide the regions in which to picture "compositional" (summing to one) data, such as rumen relative abundances.

For both the heatmap and simplex plot, the bacterial taxa were reduced to four "rumenotypes" or clusters using *k*-means clustering; distance between taxa were established using the relative abundance profiles across animals. Four taxa clusters were chosen for two reasons: first they capture the bulk of the variation across taxa and second, this choice is the largest that can be readily pictured in simplex plot. It is important to stress that the rumenotypes cluster by co-occurrence across animals, and not by phylogenetic relationship of the bacterial taxa. Relative abundances of taxa in a cluster were summed and assigned to the four rumenotypes. For the heatmap, rows in the resulting 520×4 matrix of animals by rumenotypes were again clustered using *k*-means. Cluster centre values were log transformed to accentuate the differences in the animal groups seen in the heatmap.

RESULTS AND DISCUSSION

The two graphical approaches are now described, both employing clustering, with the first using heatmaps and the second using simplex plots. Each has advantages; the simplicity of the heatmap reveals the animal clusters readily whereas the greater detail of the simplex plot (which does not cluster animals) reveals the relationship between animal groups and methane emission more clearly.

Heatmaps. The resulting dendrogram of rumenotype clusters is shown against the columns, and the dendrogram of animal similarities by rumen bacterial community is shown on the left, in Figure 1. Finally, zero-centred breeding values for CH_4/DMI were added in the right column of Figure 1. Three aspects of this heatmap are noteworthy. First, the row profiles (animals) are broadly of two types, indicated by the row dendrogram. The row microbiome profiles run either "red, red, yellow, yellow" or "red, yellow, red, yellow". Second, the methane breeding values run counter to this division. The two lowest values, –0.037 and –0.016, lie in the lower and upper microbiome profile types. This suggests two broad bacterial community types, each with high and low methane variants. This will be seen again, in different form, in the second graphical presentation. Third, the relative abundance of Rumenotype4 is generally high.

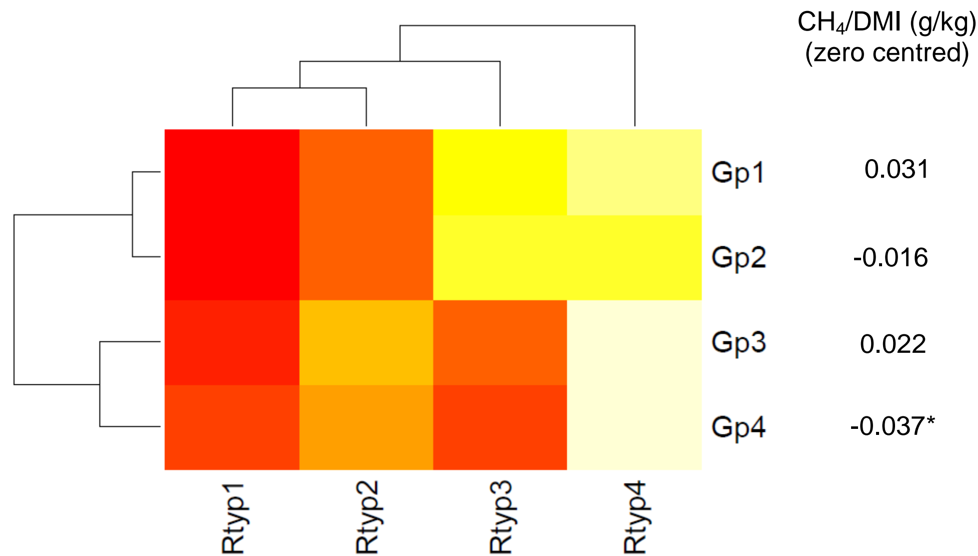


Figure 1. A heatmap (red low, yellow high) of relative abundance of bacterial clusters within the sheep microbiome. Four animal clusters (Gp1-Gp4, with animal counts of 70, 95, 214 and 141 respectively) by four microbiome clusters (Rumenotype1-Rumenotype4) are shown. Zero-centred breeding values for CH_4/DMI for each animal group (with a more negative breeding value corresponding to lower expected CH_4/DMI in progeny) are shown in the right hand column. The main conclusion is that high abundance of Rumenotype4 and moderately low abundance of Rumenotype3 can be associated with lower methane emission (the asterisked Gp4 animal row).

Simplex plots. Using the same co-occurring taxon clusters (Rumenotype1-Rumenotype4) but without animal clustering, we can plot all 520 sample relative abundance profiles in a three-dimensional simplex (a tetrahedron), colouring the resulting points according to level of methane emission. The resulting plot can be rotated; a selection of views is given in Figure 2. It suffices to plot the first three components of each point; this corresponds to mapping the tetrahedron spanned by the unit vectors in R^4 onto the unit vectors in R^3 together with the origin, in the canonical way.

Three main conclusions can be drawn from Figure 2. First, the dimensionality of the clustered microbiome is essentially two (seen in this approach in the planarity of the points). Second, the high relative abundance of Rumenotype4 (the (0,0,1) corner where many points lie). Third, lower CH_4/DMI (coloured red) occur in a continuum across the microbiome space, with some concentration at high levels of Rumenotype4.

Conclusions. We conclude with some overall remarks:

i) The microbiomes are essentially two-dimensional (this emerges from both graphical approaches). This is justified generally in that the four rumenotype clusters used capture the bulk of across animal variation.

ii) The relationship of these dimensions to methane emission is not yet clear cut using this data set; each graphic indicates that low emission occurs across the spectrum of animal groups.

iii) Each graphical approach offers advantages. For example, the heatmap shows the relative abundances in the animal groups more clearly while the simplex plot shows the two-dimensionality of the microbiome space more clearly.

iv) The challenge remaining is to provide a fundamental biological explanation for the core two-dimensionality of the rumen microbiome suggested here.

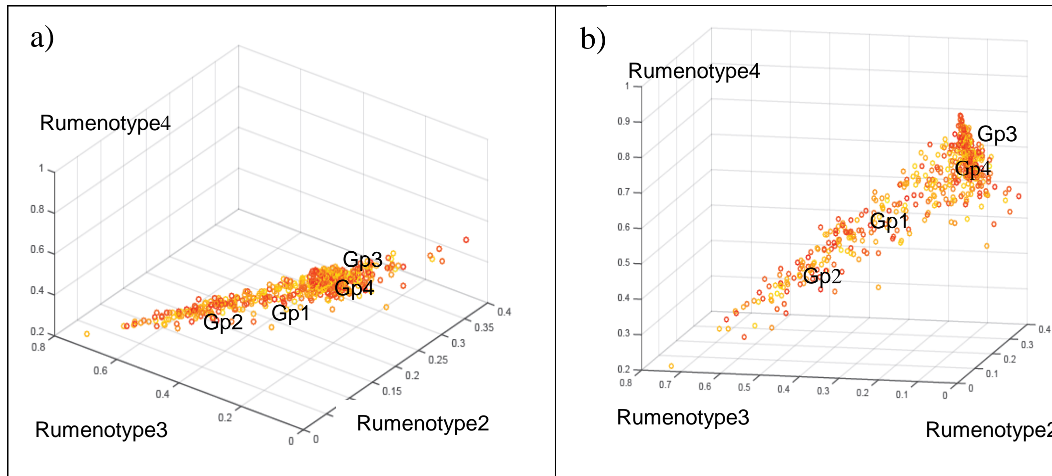


Figure 2. Tetrahedron plot of animal microbiome profiles. Each point represents the four-component (summing to one) relative abundances of Rumenotype1-Rumenotype4 for an animal microbiome sample. These are plotted within a three-dimensional tetrahedron, spanned by (0,0,0), (1,0,0), (0,1,0) and (0,0,1) ; colours represent level of CH_4/DMI (with red low and yellow high, but here note that these refer to level of methane emission, not relative abundance as in Figure 1). Four vertices of the unit cube represent extreme microbiomes (specifically, (0,0,0), (1,0,0), (0,1,0) and (0,0,1) correspond to all weight on Rumenotype1, Rumenotype2, Rumenotype3 and Rumenotype4 respectively) . Two perspectives are shown (the source object can be rotated in Matlab): in a) the planar nature is apparent while in b), the Rumenotype4 animals, with red (low) colour, show a microbiome region of low methane emission.

ACKNOWLEDGEMENTS

This study was funded by the New Zealand Pastoral Greenhouse Gas Research Consortium (PGgRc) and the New Zealand Agricultural Greenhouse Gas Research Centre (NZAGRC). Beef and Lamb New Zealand, AgResearch Limited, Lincoln University and On Farm Research are acknowledged for generating the animals and providing access to the Central Progeny Test flocks.

REFERENCES

Gilmour, A.R., Gogel, B.J., Cullis, B.R. and Thompson, R. (2009) ASReml User Guide Release 3.0. VSN International Ltd.

Greenacre, M. *Theory and Applications of Correspondence Analysis*. London: Academic Press (1983).

Kittelmann S., Seedorf H., Walters W.A. *et al.* (2013) *PLoS One* **8**, e47879.

Kittelmann S., Pinares-Patiño C.S., Seedorf H. *et al.* (2014) *PLoS One* **9**, e103171.

Pinares-Patiño C.S., McEwan J.C., Dodds K.G. *et al.* (2011) *Animal Feed Science and Technology* **166-167**:210-218.

Pinares-Patiño C.S., Hickey S.M., Young E.A. *et al.* (2013) *Animal* **7**: suppl 2: 316.

McDonald D., Price M.N., Goodrich J. *et al.* (2012) *The ISME Journal* **6**: 610.

Ross, E.M., Moate, P.J., Marett, L.C. *et al.* (2013) *PLoS One*, **8**, e73056.

Wallace R.J., Rooke J.A., Duthie C.A. *et al.* (2014) *Scientific Reports* **4**, 5892.

Proc. Assoc. Advmt. Breed. Genet. 21: 89-92

PREDICTION OF RESIDUAL FEED INTAKE FROM GENOME AND METAGENOME PROFILES IN FIRST LACTATION HOLSTEIN-FRIESIAN DAIRY CATTLE

M. Wang[1], J. E. Pryce[1,2], Keith Savin[1] and B. J. Hayes[1,2]

[1]Biosciences Research Division, Department of Economic Development, Agribio, 5 Ring Road, Bundoora, VIC 3086, Australia
[2]La Trobe University, Bundoora, VIC 3086, Australia

SUMMARY

Traits such as feed efficiency in dairy cattle are likely to be influenced by the genome of the host and the composition and abundance of microbiomes in the rumen. Here we describe an integrative approach that utilizes both genomic (SNP) and rumen microbiome data to predict future residual feed intake (RFI). The approach was tested in a small sample, of 28 Australian Holstein-Friesian dairy cattle that had 30K SNP genomic predictions for RFI and rumen microbiome profiles. The genomic and microbiome profile predictions were combined using a linear regression model. Results are very preliminary due to the small size of the data set, however, the prediction accuracy in cross validation was maximized when both SNP and rumen microbiome profiles were used (r=0.57; 95% CI: 0.33:0.72). These results, while promising, should be repeated in a larger data set.

INTRODUCTION

Feed efficiency is a key economic trait for livestock species, including dairy cattle. One measure of feed efficiency is residual feed intake, which is the approximate difference between the actual feed intake and estimated feed intake based on a regression model that takes into account energy costs for body maintenance and production over a defined production period (Connor, 2014). Macdonald *et al.* (2014) and Pryce *et al.* (2014) both demonstrated that genomic estimated breeding values (GEBV) for RFI could be derived which predict residual feed intake (RFI) with moderate accuracy. In addition to the cow's own genome, the profile of the rumen microbiome (species composition and abundance) has been shown to be associated with some traits, particularly methane emissions (Ross *et al.* 2013a; Kittelmann *et al.* 2014). So an obvious question is, can we improve predictions of future RFI phenotypes by integrating genomic predictions from SNP genotypes with rumen microbiome profiles. This seems promising, as integration of genomic, transcriptomes, proteomics and metabolomics information has already returned high accuracy in predicting type 2 diabetes (Chen *et al.* 2012). The objectives of this study were to investigate: (1) can rumen microbiome profiles be used to predict RFI for dairy cattle? (2) can the accuracy of prediction be increased by integrating using GEBV and rumen microbiome profiles?

MATERIALS AND METHODS

The dataset included 28 first parity Australian Holstein-Friesian dairy cows which were born in 2 different years at the Ellinbank research station, Victoria, Australia. Fifteen out of 28 cattle were born in July to September 2008, referred to here as FCE1 animals. The rest, referred as FCE2 animals, were born in July to September 2009. Rumen samples and dry matter intake data were collected during 1st lactation, which was in February 2011, at the age of 938 ± 12 day for FCE1 cattle, and in November 2011, at the age of 812 ± 18 day for FCE2 cattle, respectively. All animals were fed similar diets, which constituted predominantly of alfalfa hay pressed into cubes. In lactating cows the diet was supplemented with crushed wheat. Feed was always available *ad libitum*. RFI phenotypes were calculated by regressing DMI on fixed effects and liveweight and

growth in heifers and DMI on fixed effects, liveweight and production in lactating cows as described by Macdonald *et al.* (2014).

To calculate GEBVs for the 28 animals, the reference dataset comprised a total of 815 Australian growing heifers of which 74 also had RFI measurements in first lactation (Macdonald *et al.* 2014). The genotype data described by de Haas *et al.* (2012) that comprised 30,949 SNP were used to construct the genomic relationship matrix using the Yang *et al.* (2010) method. The analysis, using G-REML, was performed using ASReml software (Gilmour *et al.* 2009). A bivariate model similar to that derived from Pryce *et al.* (2015) was fitted, so that the covariance between growing heifer and cow RFI could be estimated. The model used was:

$$y_T = X_T b_T + Z_T g_T + e_T$$

Where y_T was the $2xn$ matrix of observations on all traits, X_T was the incidence matrix for fixed effects, b_T was the matrix of solution of fixed effects, Z_T was an incidence matrix mapping records to animals, g_T was the corresponding genomic breeding values for animals with genotypes for all traits, and e_T was a $2xn$ matrix of residual terms. The g_T was assumed to be distributed as $N(0, G \otimes K)$, where G was the animal by animal genomic relationship matrix and K was a $2x2$ matrix of additive genetic variances between heifers and cows. Then $V(e_T) = R \otimes I$, where R was a $2x2$ matrix of error variances and I was an nxn identify matrix.

Twenty-eight microbiome samples were extracted using the PowerMaxSoil DNA Isolation kit (MoBio) and sequenced on the HiSeq 2000 (Illumina) as per Ross *et al.* (2013b). Raw sequencing reads were trimmed from 5'-end and retained for downstream analysis if the 5'-end reached a maximum of 3 bases whose phred quality score were <15; the average remaining read quality was ≥20; and remaining read length was ≥50bp. This resulted in more than 268 million reads from all samples passed filtering. Trimmed reads were subsequently aligned to reference library using Bowtie2 (version 2.2.2; Langmead and Salzberg 2012). The reference library was composed of assembled rumen microbiome contigs from 3 smaller collections of sequences (Hess *et al.* 2011; Ross *et al.* 2012; Ross *et al.* 2013b). Contigs from the 3 sources were concatenated and sequences <250bp were removed. An overall alignment rate of 17.36% from all animals was attained and the distribution of sequences aligning to reference contigs was plotted in Figure 1.

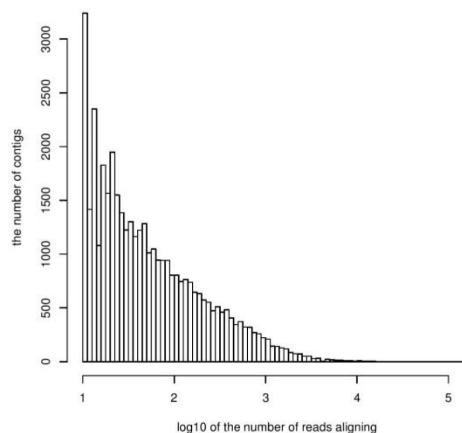


Figure 1. A histogram of read distribution. The majority of contigs had 10 to 100 reads aligned to. Contigs that had less than 10 reads aligning were removed from analysis.

Rumen microbiome profile prediction (RMP) for RFI was performed in the free R statistical software (version 3.1.2; The R Foundation for Statistical Computing; http://www.r-project.org/)

and package rrBLUP (Endelman *et al.* 2011) was used. A metagenomics relationship matrix calculated as per Ross *et al.* (2013a) was fitted into best linear regression model (BLUP) and validated using two-fold cross-validation (CV), where FCE1 and FCE2 were either training or validation sets and an alternative procedure, called leave-one-out where we sequentially removed just one animal from the dataset to estimate its genomic breeding value using the remaining data. Animals being predicted were always omitted from training set. Integrative (genomic and metagenomics) prediction was performed in R statistical software. Twenty-eight measured RFI values, GEBVs for RFI and RMP were fitted into a linear regression model. The coefficients in the output were multiplied with GEBV and RMP respectively to calculate the integrative predicted RFI. Accuracy was assessed by Pearson's correlation, 'r', that is, the correlation between the measured values with predicted values. Ninety-five percent confidence interval (CI) was calculated via bootstrapping with 10,000 replicates. Coding scripts are available upon request.

RESULTS AND DISCUSSION

The accuracy of genomic prediction was 0.33 (95% CI: 0.07:0.59; Table 1). A non-zero accuracy was observed for RFI calculated using rumen microbiome profile prediction under leave-one-out CV (r=0.49; 95% CI: 0.2:0.67; Table 1), but the accuracy of rumen microbiome profile prediction under two-fold CV was much lower (r= 0.08; 95% CI: -0.39:0.34; Table 1).

When both the cow's genome and rumen microbiome information were used for predicting RFI, the accuracies were the highest in both two-fold (r=0.38; 95% CI: 0.05:0.65; Table 1) and leave-one-out (r=0.57; 95% CI: 0.33:0.72; Table 1) testings.

Table 1 accuracy comparison among genomic, metagenomics and integrative predictions

Sequence source	CV[2] method	Correlation	95% CI[#]	Significant
Cow's Genome	Not available	0.33	(0.07, 0.59)	Y
Rumen microbiome	Two-fold	0.08	(-0.39, 0.34)	N
	Leave-one-out	0.49	(0.2, 0.67)	Y
Integration[1]	Two-fold	0.38	(0.05, 0.65)	Y
	Leave-one-out	0.57	(0.33, 0.72)	Y

[1]Integration: both cow's genome and rumen microbiome information were used.
[2]CV: cross validation.
[#]95% Confidence interval of the Pearson's correlation coefficient r based on 10,000 bootstraps.

Our results showed two main findings: firstly, rumen microbiome profiles may be able to predict RFI in some circumstances; secondly, integrating genomic and metagenomics information can increase prediction accuracy. The idea of integrating genetic information has already been realised in human research (Chen *et al.* 2012), but to our knowledge this study is the first to apply it to predict RFI in livestock. Four main elements affect the performance of prediction from rumen microbiome profiles: the number of samples in the study, size of the reference library, diversity of reference library and sequence depth (Ross *et al*. 2012). Even though the number of samples involved in our study was small, by updating the rumen reference library and maintaining a sequence depth of a minimum of 3 million reads, we obtained a similar accuracy as that in Ross *et al.* (2013a) study. We saw a growth of overall alignment rate as compared with Ross *et al.* (2013a). This could continually be improved by adding internationally collaborative references such as the Hungate1000 database (Nordberg *et al.* 2014). Currently rumen samples are still relatively hard to obtain; therefore a wider mining of ruminant metagenomics sequencing data will rely on technical improvements on sample collection.

In conclusion, microbiome information appears to be useful in predicting RFI of the same host animals. Prediction accuracy could be increased when both cow's genome and rumen microbiome

profiles are used together, though given the small samples size used here, the analysis needs to be repeated in a larger data set.

REFERENCES

Chen R., Mias G.I., Li-Pook-Than J., Jiang L., Lam H.Y., *et al* (2012) *Cell* **148**: 1293.

Connor E.E. (2014) *Animal*. **8**:1.

de Haas Y., Calus M. P., Veerkamp R.F., Wall E., Coffey M.P., *et al* (2012) *J. Dairy Sci.* **95**: 6103.

Endelman J.B. (2011) *Plant Genome* **4**: 250.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) 'ASReml User Guide' Release 3.0, VSN International Ltd.

Kittelmann S., Pinares-Patino C.S., Seedorf H., Kirk M.R., Ganesh S., *et al* (2014) *PloS one* **9**: e103171.

Langmead B., Salzberg S. (2012) *Nat. Methods* **9**:357.

Macdonald K.A., Pryce J.E., Spelman R.J., Davis S.R., Wales W.J., *et al* (2014) *J. Dairy Sci.* **97**: 1427.

Nordberg H., Cantor M., Dusheyko S., Hua S., Poliakov A., *et al* (2014) *Nucleic Acids Res.* **42**: D26.

Pryce J.E., Gonzalez-Recio O., Nieuwhof G., Wales W., Coffey M.P., *et al* (2015) *J. Dairy Sci.* (submitted).

Pryce J.E., Wales W.J., de Haas Y., Veerkamp R.F. and Hayes B.J. (2014) *Animal.* **8**(1): 1.

Ross E.M., Moate P.J., Marett L.C., Cocks B.G. and Hayes B.J. (2013a) *PloS one* **8**(9): e73056.

Ross E.M., Moate P.J., Marett L., Cocks B.G. and Hayes B.J. (2013b) *J. Dairy Sci.* **96**(9): 6030.

Ross, E.M., Moate P.J., Bath C.R., Davidson S.E., Sawbridge T.I., *et al* (2012) *BMC Genetics* **13**: 53.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., *et al* (2010) *Nat. Genetics* **42**:565.

Proc. Assoc. Advmt. Breed. Genet. 21: 93-96

ASSOCIATIONS OF RUMEN VOLATILE FATTY ACIDS WITH PHENOTYPIC AND GENETIC VARIATION IN METHANE PRODUCTION TRAITS IN ANGUS CATTLE

H.C. Smith[1], R.M. Herd[1], K.A. Donoghue[2], T. Bird-Gardiner[2], P.F. Arthur[3] and R.F. Hegarty[4]

[1]NSW Department of Primary Industries, Beef Industry Centre, Armidale, NSW 2351
[2]NSW Department of Primary Industries, Agricultural Research Centre, Trangie, NSW, 2823
[3]NSW Department of Primary Industries, Agricultural Institute, Menangle, NSW, 2568
[4]Environmental and Rural Science, University of New England, Armidale, NSW 2351

SUMMARY

Methane emissions from beef cattle contribute to greenhouse gas emissions in the atmosphere and waste gross feed energy consumed by cattle. Screening tests for methane emissions in cattle would be useful in genetic selection programs to reduce emissions. This paper reports results for 136 yearling-age Angus heifers and bulls tested for methane production in respiration chambers, and rumen fluid samples taken 3 hours post-feeding analysed for concentrations of volatile fatty acids (VFAs). A subsample of animals had repeat rumen samples taken 24 hours after feeding. The animals were fed a roughage ration offered at 1.2-times maintenance through testing. Concentrations of major VFAs (acetate, propionate and butyrate) and their proportions in the 3 hours post-feeding sample were strongly associated with methane production (g/d) (correlation coefficients up to 0.62), but less strongly with methane yield and residual methane production (correlation coefficients up to 0.17 and 0.28, respectively). Taking a rumen fluid sample during peak fermentation revealed stronger associations between methane emissions and VFA concentrations than previously reported for samples collected 24 hours after feeding. These relationships open the possibility of using VFA concentrations in rumen samples obtained at peak fermentation as indicator traits for methane emissions. For genetic selection, scrutiny of VFA as a screening test for methane emissions is still warranted.

INTRODUCTION

Cattle and sheep emit methane, a potent greenhouse gas, as part of the fermentation of feed in their rumen. There exists phenotypic and genetic variation in methane production traits of sufficient magnitude in Angus cattle that breeding for cattle with lower emissions is possible (Bird-Gardiner *et al.* 2015; Donoghue *et al.* 2015). From mechanistic fermentation models, changes in methane production should correspond with changes in the supply of hydrogen from the formation of volatile fatty acids (VFA) during diet substrate fermentation (Ellis *et al.* 2008). However Herd *et al.* (2013) were not able to detect significant associations between methane production rate (L/day) (MP; L/day) and concentrations of VFA in rumen fluid (mmoles/L) collected 24 hours after feeding, and only modest correlations with methane yield (MY = MP per unit feed intake; L/kgDMI) were observed. Those authors concluded that phenotypic associations were too low for VFA concentrations at 24 hours after feeding to be used in screening for high or low methane emitting cattle.

This experiment investigated whether VFA concentrations in rumen fluid collected shortly after feeding, during peak fermentation, were correlated with phenotypic and genetic variation in methane production traits and offer a strategy for screening cattle for methane emissions.

MATERIALS AND METHODS

A total of 140 animals (62 heifers and 78 bulls) born in the NSW Department of Primary Industries Trangie Angus research herd (Donoghue *et al.* 2015) in 2013 were tested for methane

production as yearlings in 2014 at the University of New England (UNE) methane measurement facility. They were sampled for rumen fluid before being measured for methane production. Animals were moved to UNE in cohorts of approximately 40 animals of the same sex, and kept in group pens for at least 3 days, or until they were tested, in groups of 10. Groups of animals were moved from group feeding pens (last meal offered 24hours earlier) to individual pens, and offered a meal containing their individual daily allowance. Animals were fed lucerne:cereal chaff at 1.2 times maintenance level, based on their body weight record before transport (TWT). Rumen samples were then taken from each animal 3 hours after feed was offered (the same day they were moved to individual pens). This was predicted to be close to peak methane production, based on Deighton *et al.* (2014). A 2-day methane production test was conducted using open circuit respiration chambers (Herd *et al.* 2014) after 2-3days of feeding in the individual pens. For 2 groups of heifers (n=20), a repeat sample was taken following the methane test period, representing approximately 24 hours after feeding, as done previously by Herd *et al.* (2013). Rumen samples, collected using stomach tubing, were preserved by acidification and stored at -18°C. Liquid chromatography was performed on samples to analyse VFA concentrations.

Dry matter intake (DMI; kg/d) was calculated as the average of measured dry matter intake for the two days of methane measurement. Methane production rate was taken as an average over the two days of measurement. Methane yield was calculated as MPR divided by DMI. Residual methane production (RMP; g/d) was calculated as the residual from actual MPR against DMI predicted MPR, from the regression of test data for MPR against DMI, as described in Herd et al. (2014). Phenotypic associations between traits were assessed using Pearson's correlation tests (R Core Team 2014). To assess the associations for VFA traits with genetic variation in methane emissions, correlations were determined with within-herd Estimated Breeding Value (EBV) for MPR, MY and RMP, the latter calculated as described in Donoghue *et al.* (2015). Four animals were removed from the analysis because of large feed refusals.

Table 1. **Summary statistics for n=136 yearling Angus bulls and heifers tested for methane production and with rumen fluid samples taken 3 hours post feeding**

Trait	Average	SD	Maximum	Minimum
Pre-test liveweight (TWT; kg)	390	58	512	270
Dry-matter intake (DMI; kg/d)	6.2	0.8	7.8	3.5
Methane production rate (MPR; g/d)	137	20	180	89
Methane yield (MY; g/kg DMI)	21.9	1.2	26.1	18.2
Residual MPR (RMP; g/d)	0.0	6.6	23.1	-18.0
Acetate (mmoles/L)	62.3	11.5	94.2	22.0
Propionate (mmoles/L)	17.8	3.8	27.4	4.0
Iso-butyrate (mmol/L)	0.24	0.10	0.43	0.02
Butyrate (mmoles/L)	11.3	4.5	26.7	2.1
Iso-valerate (mmol/L)	0.94	0.33	1.75	0.20
Valerate (mmol/L)	1.52	0.52	3.31	0.27
Total VFA (mmoles/L)	94.1	16.1	136.8	29.3
Acetate% (% total VFA)	66.3	4.9	75.0	57.2
Propionate (% total VFA)%	18.9	2.4	25.6	13.7
Butyrate% (% total VFA)	12.0	4.2	23.4	6.3
EBV_MPR (g/d)	-0.2	4.1	8.4	-12.8
EBV_MY (g/kg DMI)	0.00	0.39	1.02	-0.74
EBV_RMP (g/d)	0.02	2.34	6.49	-4.40

RESULTS

Summary statistics for all animals with 3-hour post-feeding VFA measurements are presented in Table 1. There was substantial variation in methane traits, with MPR strongly influenced by TWT and DMI (Table 2). However, these relationships did not persist for MY or RMP, and there was variation independent of feed intake (CV for MY and RMP; 5.2% and 3.5%). The concentrations and proportions of VFAs (except total VFA) had significant correlations with MPR, but most not with either MY or RMP. Propionate concentration was the only VFA trait which had a significant correlation with MY and iso-valerate was the only VFA with a significant correlation with RMP. Analysis of variance showed that TWT and sex significantly influenced the VFA traits.

Phenotypic variation in the VFA concentrations and ratios were not associated with genetic variation in the 3 methane traits, as indicted by a lack of significant correlation coefficients with the EBV for MPR, MY or RMP (Table 2). The exception was iso-valerate concentration (a minor VFA) which had a significant (P<0.05) correlation with the EBV for MY and RMP (regression coefficients; 0.18 (±0.07) and 0.03 (±0.01)).

Table 2. Pearson correlations for methane production rate (MPR), methane yield (MY), residual methane production (RMP) and their respective within-herd EBV with pre-test animal weight (TWT), dry-matter intake (DMI) and rumen volatile fatty acid concentrations

	MPR (g/d)	MY (g/kg DMI)	RMP (g/d)	EBV_MPR (g/d)	EBV_MY (g/kg DMI)	EBV_RMP (g/d)
TWT (kg)	0.93***	0.13	-0.01	0.35***	0.03	0.03
DMI (kg/d)	0.94***	0.04	-0.06	0.32***	0.02	0.02
Acetate (mmoles/L)	-0.26**	-0.13	-0.03	-0.07	-0.06	-0.06
Propionate (mmoles/L)	0.19*	-0.17*	-0.13	-0.00	-0.04	-0.04
Iso-butyrate (mmol/L)	-0.55***	-0.12	0.08	-0.08	-0.02	-0.03
Butyrate (mmoles/L)	0.46***	0.05	-0.07	-0.03	-0.07	-0.07
Iso-valerate (mmol/L)	-0.21*	0.13	0.28***	0.13	0.22**	0.21*
Valerate (mmol/L)	-0.24**	-0.07	0.06	-0.03	-0.05	-0.05
Total VFA (mmoles/L)	-0.03	-0.12	-0.06	-0.06	-0.07	-0.07
Acetate%	-0.62***	-0.05	0.06	0.00	0.00	0.01
Propionate%	0.39***	-0.13	-0.12	0.02	0.02	0.02
Butyrate %	0.57***	0.12	-0.04	-0.04	-0.04	-0.04

*P<0.05; **P<0.01; ***P<0.001.

Linear models to predict methane from multiple VFAs indicated significant (P<0.05) models utilising the major VFAs and Isovalerate, with or without Iso-butyrate, and some interactions, explain variation in methane traits and their EBVs. These models were stronger for methane traits (P<0.001 and R^2=0.52,0.28 and 0.23 for MPR, MY and RMP), and weaker for EBVs (P<0.05, R^2=0.14 for EBV_MPR and P<0.01, R^2=0.19 and 0.17 for EBV_MY and EBV_RMP).

For the 19 animals with repeated rumen fluid samples and methane data, VFA concentrations were generally higher in samples taken at 3 hours compared to those taken 24 hours after feeding (eg. Total VFA 96.9mmoles/L at 3hr and 61.4 mmoles/L at 24hr). Correlations between the 3-hour and 24-hour sample values were generally low or negative, except for propionate concentration at -0.47 (only significant correlation, P<0.05), with low correlations for their proportions.

In general, the magnitude of the correlations for the major VFAs with MPR were higher for the 3-hour sample (0.26 to 0.13) than for the 24-hour sample (0.05 to 0.07). Correlations at 3-hours and 24-hours with MY and RMP were in different directions. However, the only statistically-

significant (P<0.05) correlation in this data set was acetate proportion at 24 hours after feeding with MY (r=0.46), with none of the correlations with 3hr samples reaching significance.

DISCUSSION

Comparing the results from the repeated samples taken 24 hours after feeding during this study with the results from Herd *et al.* (2013), all taken 24hours after feeding, the concentrations of the major fatty acids fall within one standard deviation of the mean from the previous study. In common with Herd *et al*. (2013), the associations between total VFA, concentrations of the three most abundant VFA, and their molar proportions in rumen fluid samples collected 24 hours after feeding, with MPR, were weak. With MY the relationships were in the same direction as reported by Herd *et al*. (2013), although mostly non-significant in the present experiment, presumably due to the small sample size. Herd *et al*. (2013) concluded that the strength of the phenotypic associations between VFA concentrations in rumen fluid taken 24 hours after feeding with methane emissions to be too low for an accurate screening test for high or low emitting cattle.

Taking a rumen fluid sample during peak fermentation revealed stronger associations between methane production and VFA concentrations than reported by Herd *et al*. (2013) for samples collected 24 hours after feeding. These stronger relationships open the possibility of using VFA concentrations in rumen samples obtained at peak fermentation as indicator traits for variation in MP if feed intake cannot be measured. The significant relationships with MY reported by Herd *et al.* (2013) were not replicated in the VFA concentration in 24-hour rumen samples collected in this study. For genetic selection, more careful scrutiny of VFA as markers for methane emissions is still warranted.

Examining the relationships between the new EBVs for methane traits and VFA concentrations, the only VFA with a significant relationship with methane trait EBVs was Iso-valerate concentration, a minor VFA, with MY and RMP and these relationships were still relatively weak. Iso-valerate is associated with fibre digestion (Liu *et al.* 2009), and differences in iso-valerate concentrations have been related to Net Feed Intake (Hernandez-Sanabria *et al.* 2010). This relationship is worth further investigation to understand how rumen fermentation changes as animals are genetically selected for increased or decreased methane emissions.

REFERENCES

Bird-Gardiner, T., K. A. Donoghue, P. F. Arthur, R. M. Herd and R. F. Hegarty (2015) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21**: (in press).

Deighton, M. H., S. R. O. Williams, M. C. Hannah, R. J. Eckard, T. M. Boland *et al.* (2014) *Anim. Feed Sci. Technol.* **197**: 47.

Donoghue, K. A., T. Bird-Gardiner, P. F. Arthur, S. H. Bird, R. M. Herd and R. F. Hegarty (2015) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21**: (in press).

Ellis, J. L., J. Dijkstra, E. Kebreab, A. Bannink, N. E. Odongo *et al.* (2008*) J. Agric. Sci.* **146**: 213.

Herd, R. M., P. F. Arthur, K. A. Donoghue, S. H. Bird, T. Bird-Gardiner and R. S. Hegarty (2014) *J. Anim. Sci.* **92(11)**: 5267.

Herd, R. M., S. H. Bird, K. A. Donoghue, P. F. Arthur and R. S. Hegarty (2013) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **20**: 286.

Hernandez-Sanabria, E., L. L. Guan, L. A. Goonewardene, M. J. Li, D. F. Mujibi *et al* (2010) *Appl. Environ. Microbiol.* **76(19)**: 6338.

Liu, Q., C. Wang, Y. X. Huang, K. H. Dong, W. Z. Yang *et al.* (2009) *J. Anim. Physiol. Anim. Nutr.* **93(6)**: 716.

R Core Team (2014) 'R: A language and environment for statistical computing' R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/

Proc. Assoc. Advmt. Breed. Genet. 21: 97-100

LIVER MICRORNA EXPRESSION REVEALED THE DIFFERENCE IN FAT METABOLISM FROM CATTLE DIVERGENTLY SELECTED FOR RESIDUAL FEED INTAKE

Y. Chen[1], W. AL-Husseini[2], I. Barchia[1], C. Gondro[2], R.M. Herd[3], and P.F. Arthur[1]

[1] NSW Department of Primary Industries, Agricultural Institute, Menangle, NSW 2568
[2] Environmental and Rural Science, University of New England, Armidale, NSW 2351
[3] NSW Department of Primary Industries, Beef Industry Centre, Armidale, NSW 2351

SUMMARY

MicroRNAs (miRNAs) are short non-coding RNAs that post-transcriptionally regulate expression of mRNAs in many biological pathways. The liver plays a critical role in the synthesis of molecules that are utilized elsewhere to support homeostasis, in converting molecules of one type to another, and in regulating energy balances. We sequenced 48 liver microRNAome from young Angus bulls divergently selected for residual feed intake (RFI). In total we obtained 135 million high quality short sequence reads and more than 93 million unique mapped sequence reads. The top 10 most abundant miRNAs families expressed in liver, represented on average 46-90% of total expressed miRNAs. We did not observe any significant miRNA expression profile for low RFI animals. However, there are distinct miRNA expression patterns separating the animals into two groups that differ significantly in P8 and rib fat thickness. This is consistent with previous finding where high RFI animals had an up-regulated AHR signalling pathway, which plays an important role in fat metabolism. This suggested that some animals have a high RFI value due to excess fat metabolism.

INTRODUCTION

MicroRNAs (miRNAs) are small (~ 22 nucleotides) non-coding RNA that regulate gene expression by targeting messenger RNA (mRNA) in a sequence-specific manner, leading to either translational repression or degradation of targeted transcripts. In animals, miRNAs target the 3'untranslated regions of mRNA through a RNA-induced silencing complex (RISC), and subject to the accuracy of the sequence complementarities, either repression of translation or cleavage of the mRNA target is achieved (Huntzinger and Izaurralde 2011). MicroRNAs are now known to repress thousands of target genes and regulate cellular processes, including cellular proliferation, differentiation and apoptosis (Meltzer 2005). The aberrant expression or alteration of miRNAs also contributes to a range of human pathologies, including diabetes and cancer (Lu *et al*. 2005).

Feed efficiency is an economically important trait in beef production and can be assessed using residual feed intake (RFI) (Archer *et al*. 1999). This is the difference between an animal's actual feed intake recorded over a test period and its expected feed intake, predicted for its body weight and growth rate (Koch *et al.* 1963). Residual feed intake estimates the feed required by an animal for a given daily weight and for the maintenance of its metabolic weight. Therefore, understanding the molecular mechanism regulating RFI will help in breeding profitable animals in agriculture. Genome wide association studies have been carried out to identify gene markers associated with RFI in beef cattle (Bolormaa *et al.* 2011) and a large proportion of SNP markers associated with RFI are not located in annotated genes in bovine genome. Gene expression studies in cattle from high and low RFI selection lines have revealed a list of differentially expressed genes with functions related to extracellular matrix growth and fat metabolism in the liver (Chen *et al.* 2011). The liver plays a critical role in the synthesis of molecules that are utilized elsewhere to support homeostasis, in converting molecules of one type to another, and in regulating energy balances. MicroRNAs (miRNAs) are important natural regulators of global gene expression. The objective

of the study is to profile the expression of miRNA in liver by next generation sequencing and identify miRNAs related to the efficiency of feed utilization in beef cattle.

MATERIALS AND METHODS

Ninety young Angus bulls resulting from approximately three generations of divergent selection for RFI were used in this study. The selection lines were established in 1993 at the Agricultural Research Centre, Trangie, NSW, Australia and the animals were used for the microarray experiment previously reported by Chen et al. (2011). In brief, bulls were born in 2005 and feed intake was measured for each animal using an automated recording system in the Beef Research Feedlot Tullimba, NSW, Australia. During the 70-day test, the animals had ad libitum access to a barley-based feedlot ration containing 12 MJ metabolizable energy per kilogram dry matter and 15–17% crude protein. Post-weaning RFI was measured using the linear regression of daily feed intake on mean metabolic mid-test weight and average daily gain. Liver biopsies were taken at the end of the RFI test from 24 animals with the lowest RFI and 24 animals with the highest RFI. Total RNA from liver was isolated using TRI Reagent (Ambion, Applied Biosystems, Austin, TX, USA) following the manufacturer's protocol and the quality and integrity of RNA was assessed with the RNA 6000 Nano Lab Chip Kit using the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA).

Small RNA libraries were constructed for each animal using 1μg total RNA with NEXTflex™ Small RNA-Seq Kit v2 (Bioo Scientific, TX, USA) following the protocols supplied by the manufacturer. The libraries were sequenced at Ramaciotti Center, University of NSW with Illumina HiSeq 2000 Sequencing System.

Sequencing data were analysed using miRanalyzer (Hackenberg *et al.* 2011). In brief, known bovine miRNAs were identified by mapping all sequence reads to known bovine miRNAs in miRBase (version 21), and reads that matched known bovine miRNAs were grouped and removed from the dataset. Reads that mapped to known miRNAs in other species were grouped as homologue miRNAs. The remaining reads were aligned to libraries of known transcripts. To identify bovine-specific novel miRNAs, the remaining sequence reads were mapped to *Bos taurus* genome (bostau6, UMD_3.1). The microRNA expressed was normalized by total mapped reads from each sample and was measured as reads per million mapped reads (RPM).

Differential expression analysis was carried out with Bioconductor package DESeq (v3.1) (Anders and Huber 2010) to identify differentially expressed miRNAs between high and low RFI animals. GenStat (V17) was used for the cluster analysis and phenotype differences between clusters were evaluated using Student's t-test.

RESULTS AND DISCUSSION

We obtained 135,042,220 high quality sequence reads and 88% of the sequence reads mapped to bovine genome. About 69% (93,336,181) of sequence reads mapped to known bovine mature miRNA (bostau6, UMD_3.1). A total of 560 known miRNAs were detected for which at least one read was observed in the dataset. Only 224 miRNAs had more than 10 reads per million across most of the samples and were used for the subsequent analysis. There were large variations of expression of miRNAs and the top 10 most expressed miRNA (bta-miR-143, bta-miR-100, bta-miR-99a-5p, bta-miR-192, bta-miR-21-5p, bta-miR-122, bta-miR-148a, bta-miR-191, bta-miR-26a, bta-miR-30a-5p) accounted for 50-to 90 % of the total miRNA expression. Bta-miR-143 is the most abundant miRNA in bovine liver and takes up on average 17% of the total miRNA expression. Our results differ from reported liver miRNA expression in negative energy balance (NEB) dairy cows (Fatima *et al.* 2014). In NEB dairy cows 53 miRNAs were expressed and the top 10 miRNA accounted for more than 95% of the miRNAome. This difference was most likely due to the different miRNA extraction and library construction methods used for sequencing. In

this study we used the total RNA for the small RNA library construction, while in NEB dairy cows, an enriched miRNA protocol was used for the RNA extraction.

To identify differentially expressed miRNA we only used 224 miRNAs that had on average more than 10 RPM. We first divided the animals into high and low RFI groups and used DESeq (Anders and Huber 2010) to identify differentially expressed miRNAs between high and low RFI animals based on a negative binomial model. No miRNA expressed differently between high and low RFI animals when the Benjamini and Hochberg corrected P value cut-off of <0.05 was applied to correct for multiple testing. Next, we carried out the cluster analysis of all animals based on miRNA profile (224 miRNAs) and top 10 miRNA only as the top 10 most expressed miRNAs accounted more than 70% of the total miRNA population. The animals were clustered in two major groups. Then we divided the animals into two groups based on cluster analysis. In Table 1 we summarize phenotypes between these two groups. The most significant difference between these two groups was in fat deposition; both P8 and rib fat were different. There was no difference in ADG, DFI and RFI.

Table 1. Trait means (± standard deviation) for two groups based on the miRNA profile

Cluster	ADG	DFI	RFI	P8	RIB	IMF%	EMA
Cluster 1	1.99 ± (0.05)	11.54 ±0.30	-0.38 ±0.24	11.77 ±0.57	9.73 ±0.60	5.03 ±0.10	82.55 ±0.68
Cluster 2	2.04 ±0.06	11.16 ±0.34	-0.52 ±0.22	10.05 ±0.54	8.29 ±0.36	4.94 ±0.10	80.38 ±0.91
P	0.570	0.404	0.671	0.033	0.048	0.565	0.063

ADG: average daily gain during the 70day; RFI test. DFI: average daily feed intake; P8: P8 fat thickness (ultrasound) at the end of RFI test; RIB: RIB fat thickness (ultrasound); EMA: eye muscle area (ultrasound); P: probability of significance.

It is believed that RFI is highly associated with the energetic costs of protein turnover and basal metabolic rate and that selection for low RFI animals will reduce maintenance energy requirements (Richardson and Herd 2004). Global miRNA profiling of the current study showed clear expression patterns related to fatness instead of RFI, although the animals were from the third generation of divergent selection lines for RFI. Indeed, there was a significant correlation between P8 fat thickness and RFI (Figure 1). It is likely that animals with high P8 fat (>1 sd) are high RFI animals and animals with low P8 fat (<1sd) are low RFI animals. Therefore, some proportion of RFI's variation can be explained by differences in body composition. Alternatively, at a constant weight and daily gain, some of the high RFI animals deposited more fat than the low RFI animals. There has been earlier reports on the genetic and phenotypic correlations between RFI and lipid metabolism (Robinson and Oddy 2004). Previous gene expression work with microarray identified up-regulated p450 induced xenobiotic signalling pathways in high RFI animals. P450 induced xenobiotic signalling pathway plays an important role in lipid metabolism. A recent study on miRNA in the liver of NEB dairy cows identified that mir-143 was down regulated in severe NEB cows (Fatima *et al.* 2014). One of the target genes of mir-143 is LRP2 which is involved in lipid metabolism

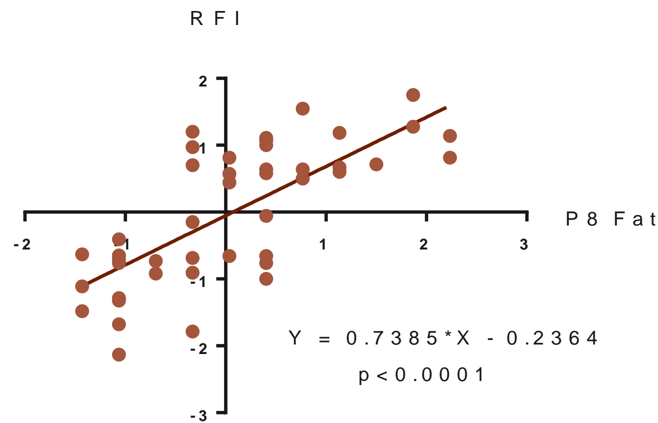


Figure 1. Scatter plot of RFI and P8 fat. The scales of X (P8) and Y (RFI) are standardized unit of standard deviation**.**

In conclusion, we have identified 224 known bovine miRNA expressed in bovine liver. There is distinct expression profile difference between high and low fatness animals. Some of the variation in RFI can be explained by the observed differences in lipid metabolism.

REFERENCES

Archer J.A., Richardson E.C., Herd R.M. & Arthur P.F. (1999) *Aust. J. Agric. Res.* **50**: 147.

Anders S. and Huber W. (2010) *Genome Biol.* **11**: R106.

Bolormaa S., Hayes B.J., Savin K., Hawken R., Barendse W., Arthur P.F., Herd R.M. and Goddard M.E. (2011) *J. Anim. Sci.* **89**: 168.

Chen Y., Gondro C., Quinn K., Herd R.M., Parnell P.F. *et al.* (2011) *Anim.Genet.* **42**: 475.

Fatima A., Lynn D., O'Boyle P., Seoighe C. and Morris D. (2014) *Bmc Genomics* **15**: 279.

Hackenberg M., Rodríguez-Ezpeleta N. and Aransay A.M. (2011) *Nucleic. Acids Res.* **39**: W132.

Huntzinger E. and Izaurralde E. (2011) *Nature Review Genetics* **12**: 99.

Koch R.M., Swiger L.A., Chambers D. and Gregory K.E. (1963) *J. Anim. Sci.* **22**: 486.

Lu J., Getz G., Miska E.A., Alvarez-Saavedra E., et al. (2005) *Nature* **435**: 834.

Meltzer P.S. (2005) *Nature* **435**: 745.

Richardson E.C. and Herd R.M. (2004) *Aust. J. Exp. Agr.* **44**: 431.

Robinson D.L. and Oddy V.H. (2004) *Livest. Prod. Sci.* **90**: 255.

Proc. Assoc. Advmt. Breed. Genet. 21: 101-105

GENETIC SELECTION FOR LITTER SIZE IN CATTLE

L. R. Piper[1], B. M. Bindon[3], A. A. Swan[2], and H. G. Brewer[1]

[1]CSIRO Agriculture Flagship, F D McMaster Laboratory, Locked Bag 1, Armidale, 2350,
[2] Animal Genetics and Breeding Unit*, University of New England, Armidale, 2351,
[3]93 Bunburra Rd, Bunburra, 4310.

SUMMARY

Efficiency of production in beef cattle is limited by their relatively low reproduction rate. In this paper, we present the results of a fourteen year breeding program aimed at increasing prolificacy in a mixed beef x dairy herd, grazing at Armidale in the New England region of NSW. The herd was established by purchasing cows with a repeated history of twinning and bulls from dams with unusually high prolificacy. For the 51foundation cows with at least one subsequent calving record, the average prolificacy was 1.11 +/- 0.05 (123 total records. For all cows born in the herd, the mean ovulation rate was 1.12 and the mean prolificacy was 1.03. The twinning rate was lower than expected based on reports from other experimental herds aimed at increasing prolificacy that have been established in the USA, New Zealand and France. The observed genetic trends for ovulation rate and prolificacy were essentially zero but were moderately positive for fertility, cow rearing ability and reproduction rate. The desired outcome of generating a positive genetic trend in reproduction rate was achieved, albeit to a limited degree, but not as a result of genetic response in ovulation rate or prolificacy, the traits under direct selection pressure.

INTRODUCTION

Efficiency of production of temperate beef cattle herds is limited by their reproduction rate (calves weaned/cow joined*)*. In NSW for example, each breeding cow weans on average only 0.85 calves per year (Wilkins, personal communication). Increasing reproduction rate can increase efficiency of production (Dickerson, 1978) but the focus of genetic improvement programs has largely been directed to increasing fertility (cows calving/cow joined) by direct selection or reducing rebreeding interval or cow rearing ability (calves weaned /calf born) rather than prolificacy (calves born/cow calving). As demonstrated by multiple ovulation and embryo transfer experiments, there are no limits to increased prolificacy from the ability of the bovine ovary to produce more than one egg per cycle or from the ability of the uterus to carry more than one foetus. The cow has four functional mammary glands and is therefore equipped to suckle more than one calf. Milk production is more than adequate in many breeds and if not, could readily be improved by selection or crossing with dairy breeds. Despite these attributes, the frequency of twinning in most breeds is less than 2% but is higher within some of the large European breeds.

As noted by Piper and Bindon (1990), renewed interest in genetic manipulation of prolificacy in cattle began in the 1970's with experimental herds selected for increased twinning rate established in France, Australia, the USA and New Zealand. These new herds were based on highly selected foundation males and females (Piper and Bindon, 1979). Comparative twinning frequencies for the foundation females before and after purchase and for their first generation daughters were summarised by Morris and Day (1986). Cows with a minimum of two sets of twins prior to purchase averaged around 14 percent of twin births in subsequent calvings, while their daughters averaged about 8 percent of twin births. These

* AGBU is a joint venture of NSW Dept. of Primary Industry and the University of New England

twinning rates are in general agreement with expectations based on the repeatability and heritability of twinning and on the approximate selection intensities applied in establishing the respective herds.

In this paper we present the final results from a long term selection experiment for increased prolificacy in a mixed beef/dairy herd grazing at Armidale in the New England region of NSW.

MATERIALS AND METHODS

Foundation Animals. CSIRO began establishing its twin selection herd in 1973/74 by purchasing cows with a repeated history of twinning (2 or more sets of twins or a set of triplets or quads) and bulls from dams with unusually high prolificacy (3-10 sets of twins, 10 bulls) or because there was a history of twinning in the bulls pedigree (2 bulls). The foundation cows and bulls came from multiple herds and documented pedigree information was almost never available. The details of the twinning history of the 65 foundation cows and 12 foundation bulls were given in Piper and Bindon (1990). For the 51foundation cows with at least one subsequent calving record, the average prolificacy was 1.11 +/- 0.05 (123 total records). By contrast, 70 unselected Hereford cows, joined with the twin herd bulls and grazing throughout the year with the selected herd, had an average litter size of 1.01 (234 records). All foundation cows and bulls were culled prior to the February 1983 joining.

Cows born in the herd. As reported by Piper and Bindon (1990), all females born in the herd up to and including the 1980 drop, were retained and given from 6 to 8 opportunities to calve (more for the earlier, less for the later drops). There was no joining in 1981 (due to drought) or in 1982 (due to a change of joining time). For the 1983 to 1986 joinings, male and female replacements were chosen on the basis of selection indexes combining information on the twinning records of their dams and grand-dams (1983 and 1984) and for the later joinings, on the ovulation rate (determined by the technique of Holland *et al.*, 1981) and twinning records of their dam and the twinning records of their grand-dams. Details of the selection procedures for replacements entering the herd for the 1983 to 1989 joinings are given in Piper and Bindon (1990).

Observations and data analysis. Ovulation rate and reproduction records for females born from 1975 to 1986 have been included in the analyses for this paper. Single trait, repeated record mixed linear models, adjusting for fixed effects were fitted using Wombat (Meyer, 2007). The fixed effects fitted included calving year (1977 to 1989 with 12 levels) and cow age (in years from 2- 10 with 9 levels). The random effects included a direct additive genetic effect fitted with the numerator relationship matrix and permanent environment of the cow. There were between 418 and 453 animals in the pedigree depending on the trait, with 36 sires for all traits. Bivariate models were fitted to estimate genetic correlations but they were poorly behaved due to the small size of the data set and are not presented in this paper. Genetic trends for each trait were estimated by taking the single trait EBVs for each cow, averaging by year of birth (1975-1986) and calculating the regression between year of birth and average EBV. The annual trends shown in Figure 1 are scaled to the genetic standard deviation for each trait. The annual trends shown in Table 1 are multiplied by 100, so they are estimates of the annual average trait change per 100 cows.

RESULTS AND DISCUSSION

The number of cows, number of records, estimated means, phenotypic variances, heritabilities(+/-se) and repeatabilities (+/-se) and genetic trends for fertility (FERT), ovulation rate (OV), prolificacy (NCB), calf survival (SURV) and reproduction rate (NCW) are given in Table 1. The mean ovulation rate was 1.12 but the mean prolificacy was 1.03 which is a

disappointing outcome given the selection pressure applied to prolificacy in assembling the foundation animals for the herd and to prolificacy and ovulation rate in the experimental herd for the 1983-1989 joinings. This outcome no doubt reflects the observed low heritability and repeatability of both traits and the fact that, due to the requirement build up numbers in the herd, there was no selection pressure applied in choosing incoming male and female replacements until the 1983 joining. From the 1983 joining onwards, selection intensities for male and female replacements averaged 12 percent and 72 percent respectively. The heritability and repeatability of fertility and reproduction rate were significantly higher than for ovulation rate or prolificacy but the means for both traits were below or about average for beef herds in NSW.

Genetic trends for all traits for animals born from 1975 to 1986 and calving from 1977 to 1989 are shown in Table 1 and Figure 1.

Table 1. Number of cows, number of records, estimated means and phenotypic variances, heritability and repeatability for each of Ovulation rate, Fertility, Prolificacy, Cow rearing ability (Calf survival), Reproduction rate and Genetic trend (Annual genetic change *100)

Trait	Cows	Records	Mean	P. Var.	Heritability	Repeatability	Trend	Sig.
FERT	380	1387	0.80	0.16	0.09(0.05)	0.27(0.03)	0.23	0.07
OV	354	1134	1.12	0.11	0.02(0.02)	0.02(0.02)	-0.02	n.s.
NCB	347	1110	1.03	0.03	0.01(0.02)	0.03(0.03)	0.02	n.s.
SURV	347	1140	0.83	0.12	0.04(0.04)	0.07(0.03)	0.10	n.s.
NCW	380	1388	0.68	0.23	0.09(0.04)	0.20(0.03)	0.47	0.01

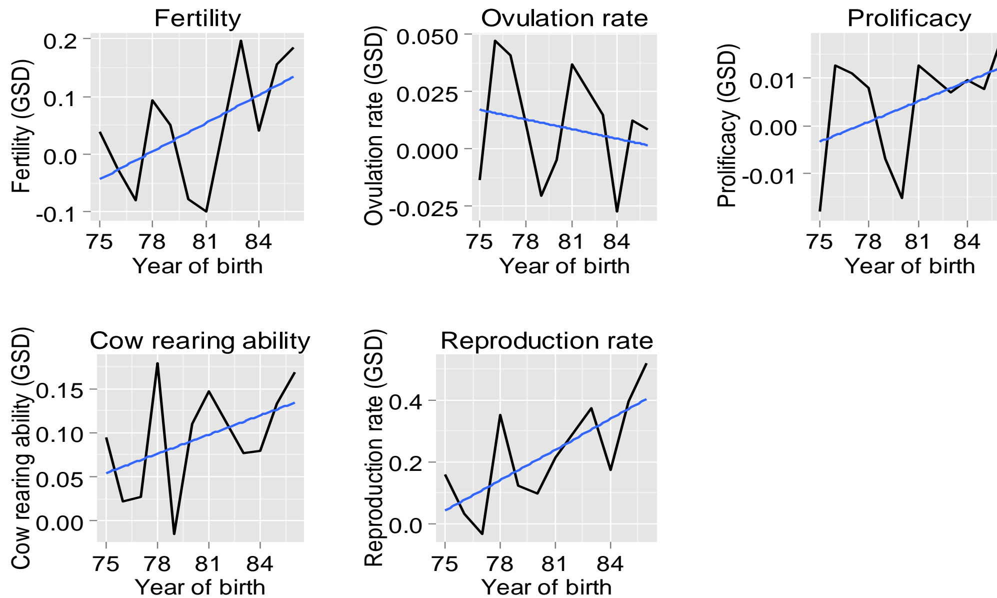


Figure 1. Genetic trends in fertility, ovulation rate, prolificacy, cow rearing ability and reproduction rate, scaled by the genetic standard deviation (GSD) for each trait.

The genetic trends for ovulation rate and prolificacy were essentially zero which is disappointing given the positive but modest selection pressure applied to both traits in choosing replacement males and females for the 1983 to 1989 joinings. By contrast the genetic trends for fertility, calf survival and reproduction rate were positive indicating slow genetic improvement in all three traits but especially in fertility and reproduction rate.

As indicated earlier, it is disappointing that the traits under direct selection pressure, ovulation rate and prolificacy, showed little or no response over the life of the experiment. Because of that observed outcome, it is not clear why there was a positive genetic trend in fertility, cow rearing ability and reproduction rate. One possible explanation may be that selection for prolificacy puts direct pressure on fertility because prolificacy cannot be observed unless the cow is pregnant. This positive genetic trend in fertility, accompanied by a positive but not significant trend in cow rearing ability has resulted in a positive but modest genetic trend in reproduction rate. The desired direction of outcome was achieved, albeit to a limited degree, but not as a result of genetic response in ovulation rate or prolificacy, the traits under direct selection pressure.

The results from this experimental herd are in sharp contrast to the results obtained in experimental herds undergoing long-term selection for increased prolificacy in New Zealand (Morris and Wheeler, 2002 and Morris, personal communication, 37 percent of twin births in 2006-2008) and in the Clay Center herd in the USA (Echternkamp *et al.*, 2002, 52% of twin births in 2000). In both these herds, as in the herd reported in this paper, the foundation animals were highly selected for repeated history of twinning. By contrast with the selection procedures employed in the present study, replacements in the Clay Centre herd in the USA, were chosen on the basis of repeated (6-8) observations of ovulation rate determined by rectal palpation (incoming young female replacements) and on the basis of repeated ovulation rate progeny tests, and in later years QTL marker adjusted EBV, for incoming replacement males.

The difference in response in prolificacy between the Clay Centre herd and our experimental herd may be due to a combination of factors including large differences in selection accuracy and intensity, the number of years that effective selection was able to be applied and in the difference in the initial response achieved in the offspring of the highly selected foundation males and females. It is also probable that there were differences in the accuracy of the records of the foundation cows and bulls. By contrast with the USDA herd, in the CSIRO herd these records were generally not documented in herd recording schemes and were based on the testimony of the producers who supplied the foundation cows and bulls. The overall response in prolificacy in the CSIRO herd may have increased had the experimental breeding program been allowed to continue. However, a decision to redirect resources to pursue non-genetic methods of increasing prolificacy in cattle resulted in the breeding program being terminated when the 1986 drop animals had their second calving in 1989.

ACKNOWLEDGEMENTS

The skilled and dedicated contributions to the conduct of this experiment during its fourteen year lifetime of Yvonne Benton, Michelle Cheers, Richard Farrell, Edward (Bud) Holland, Robert Nethery, Grant Uphill and Ross McCrae are gratefully acknowledged.

REFERENCES

Dickerson, G.E. (1978) *Anim. Prod*. **27**:367.

Echternkamp, S.E., Thallman, R.M., Kappes, S.M. and Gregory, K.E. (2002) *Proc. 7th Wld. Congr. Genet. Appld. Livest Prod., Session 08 Reproduction, Communication No 08-12*.

Holland, E.J., Bindon, B.M., Piper, L.R. Thimonier, J, Cornish, K.A. *et al.* (1981) *Reproduction Science* **4**:127.

Meyer, K. (2007) *Journal of Zhejiang University. Science. B* 8 (11): 815–21. doi:10.1631/jzus.2007.B0815.

Morris, C.A. and Day, A.M. (1986) *Proc 3rd Wld. Congr. Genet. Appld. Livest. Prod.* **XI**: 14.

Morris, C.A. and Wheeler, M. (2002) *N.Z. J. Agric Res.* **45**: 17.

Piper, L.R. and Bindon, B.M. (1979) *Proc. NZ Soc. Anim. Prod* **39**: 224.

Piper, L.R. and Bindon, B.M. (1990) *Proc. Assoc. Advmt. Anim. Breed. Genet.***8**: 381.

TOWARDS A DATA SET TO INVESTIGATE FEED EFFICIENCY IN NEW ZEALAND MATERNAL SHEEP

P.L. Johnson, S.P. Miller and K. Knowler

AgResearch, Invermay Research Centre, PB 50034, Mosgiel 9053, New Zealand

SUMMARY

Residual Feed Intake (RFI) as a measure of feed efficiency has not been reported for New Zealand maternal sheep breeds. This study reports on a pilot study that generated RFI data on 37 16-month old maternal composite ewes. Records on ten were obtained utilising a prototype automated feeder, with the remainder fed in individual pens. The animals were introduced to lucerne pellets before daily feed measurements were taken over 42 days with animals weighed twice weekly. The RFI model fitted, to determine the relationship between the liveweight and average daily weight gain and the animal intake, had an R^2 of 0.79, with the partial R^2 for liveweight the most significant at 0.70. The observed phenotypic standard deviation of RFI was 209 g of DM/d which is 8% of the average daily intake. The animals were ranked for RFI, with the 16% most efficient animals (low RFI) consuming on average 0.6kg/day less feed, or 20% less than the 16% least efficient animals (high RFI). Additional data, collected for the animals using the prototype automated feeder, included the number and size of feeding events per day which showed consistent variation. Further animals will be evaluated over the coming years with the aim to collect data on 1000 animals. The animals will be sourced from the Central Progeny Test and will represent NZ maternal sheep breeds which will be measured for a range of other production traits allowing the heritability of the trait and its genetic correlation with other traits to be estimated.

INTRODUCTION

The trait of Residual Feed Intake (RFI) proposed by Koch *et al.* (1963) as a measure of feed efficiency has been shown to be heritable in beef and dairy cattle. A meta-analysis of 39 published RFI papers by Berry and Crowley (2013) for cattle resulted in a pooled heritability estimate of 0.33 ± 0.01 (range of 0.07 to 0.62). There are very few published estimates for measures of feed intake and efficiency in sheep. Heritability estimates of 0.32 to 0.41 were reported by Forgarty *et al.* (2006) for feed intake at pasture and Cammack *et al.* (2005) reported heritability estimates of 0.11 to 0.33 for measures of feed efficiency in growing terminal sired lambs. There are currently no published genetic parameter estimates for New Zealand maternal sheep breeds.

Other important aspects of the genetics of RFI include the repeatability of the trait between growing and mature animals, and the genetic correlation with other economically important traits such as reproduction. The trait of RFI has been shown to be highly genetically correlated in cattle when measured in young growing animals and older mature animals (Herd *et al.* (2003)). Although relatively few significant genetic correlations have been observed in cattle, there is some evidence of a negative genetic correlation between RFI and puberty onset and post-partum anoestrus period intervals resulting in a delay for both (Crowley *et al.* 2011).

A study to generate RFI data over several years to estimate its heritability, repeatability and genetic correlation with other traits is due to commence in July 2015. Pilot studies have been conducted to gain insight in to the phenotypic variability of the trait, and the repeatability of the trait on a group of ewes fed different feeds (fresh cut grass versus lucerne pellets) at different times (9 and 16 months). This paper reports results based on 37 animals from the second study when the animals were 16 months old and were all fed lucerne pellets, with ten measured via the prototype automated feeder and the remainder fed in individual pens.

MATERIALS AND METHODS

Permission for this study was granted from the AgResearch Invermay Ethics Committee (Ethics Numbers 13257 and 13456). Thirty-seven 16 month old maternal composite ewes (ewes that were surplus to requirement from the Central Progeny Test and were therefore from a variety of breeds as described by McLean *et al.* 2009) previously used in a feed intake study (Study 1) as nine month olds were used. The ewes were introduced to Lucerne pellets (sourced from Dunstan Feeds, Hamilton New Zealand; Dry Matter Content 85%; Metabolisable Energy (ME) content 10.1 MJ ME) over a two-week period before the study with *ad libitum* Lucerne pellets available. A random sub-set of ten of the ewes were placed in a pen with the prototype automated feeder, with the remainder placed in neighbouring individual pens in a raised-floor shed. The feeder was designed by AgResearch and utilized a feed trough on load cells with an automated feed delivery through an auger For the ewes utilising the prototype automated feeder approximately 2.5kg of feed was always available, allowing *ad libitum* access to feed, with the weight of feed consumed recorded in real time against the animal through the use of electronic identification. The resulting data was summed across a day for an animal to provide the total feed consumed, but the number of feeding events and the average weight of feed consumed at each feeding event was also calculated. The animals in the individual pens were offered 4-5kg of feed per day at approximately 9am each morning, with the residual feed weighed 24 hours later, at least 10% residual was targeted to ensure that the animal had *ad libitum* access to feed.

The animals were weighed twice weekly, at approximately 9am. The animals were fed for forty-two days. The importance of using multiple measures of liveweight across the duration of the study to accurately estimate average daily live weight gain (ADG) was demonstrated by Johnson *et al.* (2015) using data from the first study using the same animals. The animals were Computed Tomography (CT) scanned at the beginning and conclusion of the study, but the images are not yet analysed.

A model based on Koch *et al.* (1963) was used to calculate Residual Feed Intake (RFI) using the General Linear Model (GLM) procedure in SAS: $y = \beta_0 + \beta_1 MMWT + \beta_2 ADG +$ Previous Feed + Feeder (Previous Feed) + ε; where y is measured feed intake calculated using the MIXED procedure in SAS fitting day as a repeated measure, β_0 = intercept, MMWT = metabolic mid-weight (mid-weight$^{0.75}$), ADG = the slope of model estimated by REG procedure in SAS (SAS Inst. Inc., Cary, NC) using the bi-weekly liveweight measurements and the day of measurement (with the first measurement made on day 0), Previous Feed=Lucerne pellet or grass in study one, Feeder=Individual pen or auto-feeder and ε = the residual which is the trait of RFI.

The animals were ranked based on their RFI values and the bottom and top 16% (n=6) assigned as being Low or High RFI respectively, with the remainder being assigned as medium. The significance of differences between the groups was assessed using the GLM procedure in SAS fitting RFI group as a fixed effect.

RESULTS AND DISCUSSION

The model fitted which included liveweight, ADG, previous feed and feeder in current trial and the intake of the animals had an R^2 of 0.79, of which the partial R^2 for liveweight was the most significant at 0.70. This value is higher than those reported in growing sheep by Redden *et al.* (2013) and Cockrum *et al.* (2013) whom reported R^2 values of between 0.45 and 0.65 using the same model. Computed Tomography images have also been collected on the animals which will be used to estimate the relative proportions of fat and lean in the animals. The addition of fat to the RFI model has been shown to improve the description of feed intake over and above liveweight and liveweight gain in cattle (Basarab *et al.* 2011), but not in sheep (Redden *et al.* (2013) and Cockrum *et al.* (2013)). The observed phenotypic standard deviation of RFI was 209g

of DM/d for RFI which is 8% of the average daily intake, a value consistent with values summarised for beef and dairy cattle by Williams *et al.* (2011).

The results from the grouping of the animals in to Low, Medium and High RFI groups is in Table 1. Liveweights and growth rates were not significantly different between the RFI groups as expected. Both RFI and daily dry matter intake were significantly different between the RFI groups, with the most efficient animals (low RFI) consuming on average 0.6kg/day less feed, or 20% less than the least efficient animals (high RFI). This level of difference is consistent with the results of the sheep studies of Redden *et al.* (2013) and Cockrum *et al.* (2013) and a dairy heifer study of Williams *et al.* (2011) whom observed differences of 17%, 30% and 20% respectively.

This study also involved the testing of a prototype automated feeder. In addition to providing data on the total weight of feed consumed in a day, the automated feeder provides information on the number of feeding events per day, and the weight consumed at each feeding event. A basic summary of the average number and average weight of individual feeding events per animal per day for the ten animals that utilised the automated feeder is provided in Figure 1. From Figure 1, there are consistent trends observed between animals, at the extremes one ewe (7018) had an average of 26 feeding events per day consuming an average of 128 grams of feed per feeding event, whereas another ewe (58) had an average of 14 feeding events per day but is consuming on average 288 grams of feed per feeding event. In the longer term study it will be interesting to determine whether these are heritable traits, and whether or not they are correlated to either RFI or other traits including methane emissions.

The longer term data collection, which will take place over the next 3-5 years, will aim to collect data on 800-1000 animals. The animals will be sourced from the Central Progeny Test which will represent NZ maternal sheep breeds. A range of other traits will also be measured on the animals some of which will be measured before entering the feed intake facility including weaning weight, onset of puberty and others will be measured post- time in the facility including mature weight and reproductive performance. The ram lamb brothers of the ewes will have been grown out and slaughtered as lambs, which will provide carcass breeding values for the sires. Repeated feed intake data will be collected on the same animals as mature ewes to investigate the genetic correlation between feed efficiency measured in a growing lamb and a mature ewe to consider whether the two measures should be considered as repeated measures of the same trait or different traits. There will also be the opportunity to investigate alternate predictor traits reviewed by Berry and Crowley (2013).

CONCLUSIONS

The results from this study suggest that the feed intake system established, is obtaining RFI phenotypic data with a co-efficient of variation of 8%, which is consistent with RFI data in other production species. The next stage is to collect sufficient data to estimate the heritability of RFI in NZ maternal sheep, and its genetic correlation with other economically important traits.

Table 1. Characteristics (average ± SE) among residual feed intake (RFI) group traits

	RFI Group			**Signif. of**
	Low (n=6)	**Medium (n=25)**	**High (n=6)**	**RFI Group**
Study Mid Weight (kg)	71.7 + 2.98	70.9 + 1.46	71.3 + 2.98	NS
Average Daily Gain (g/day)	231 + 24.1	248 + 11.8	247 + 24.1	NS
Dry Matter Intake/Day (kg)	2.7 + 0.18^{a}	3.0 + 0.09^{ab}	3.3 + 0.18^{b}	P<0.001
Residual Feed Intake (g/day)	-309 + 43.0^{a}	3 + 21.0^{b}	323 + 43.0^{c}	P<0.001

[1]Values within a row with different superscripts are significantly different (P<0.05)

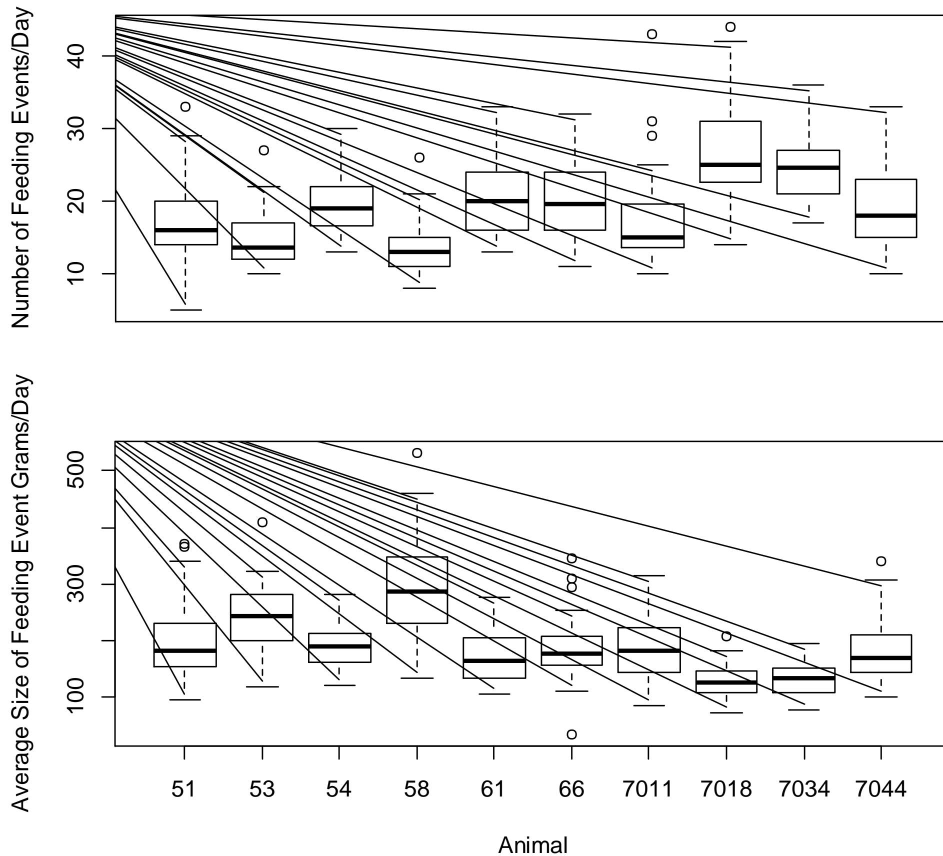


Figure 1. Boxplot summary of number of, and average size of feeding events per day for individual animal data collected over 42 days from a prototype automated feeder.

ACKNOWLEDGMENTS

This work was funded by Beef + Lamb New Zealand Genetics. AgResearch Capital Expenditure funding providing funding for the prototype automatic-feeder.

REFERENCES

Basarab, J.A.; Colazo, M.G.; Ambrose, *et al.* (2011). *Can. J. Ani. Sci.* **91:** 573.
Cammack, K. M., K. A. Leymaster, T. G. Jenkins, *et al.* (2005) *J. Ani. Sci.* **83:** 777.
Cockrum, R. R., R. H. Stobart, S. L. Lake *et al.* (2013) *Small Rum. Res.* **113:** 313.
Crowley, J. J., R. D. Evans, N. Mc Hugh, *et al.* (2011) *J. Ani. Sci.* **89:** 3372.
Herd, R. M., J. A. Archer, and P. F. Arthur. (2003) *J. Ani. Sci.* **81(13 suppl 1):** E9-E17.
Johnson, P.L.; Miller, S.P.; Knowler, K.; *et al.* (2015) *Proc. NZ Soc. Ani. Prod.* **75:** 225.
Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. (1963). *J. Ani. Sci.* **22:** 486.
McLean, N.J.; Jopson, N.B.; Campbell, A.W.; *et al.* (2006) *Proc. NZ Soc. Ani. Prod.* **66:** 368.
Redden, R.R.; Surber, L.M.; Grover, A.V. & Knott, R.W. (2013). *Small Rum. Res.* **114:** 214.
SAS. (2004). *Sas/stat 9.1 user's guide*. SAS Publishing.
Williams, Y.J.; Pryce, J.E.; Grainger, C.; *et al.* (2011) J. Dairy Sci. **94:**4715.

REPEATABILITIES FOR METHANE EMISSION IN MERINO EWES ON PASTURE ACROSS DIFFERENT AGES

S. Dominik[1] and V.H. Oddy[2]

[1]CSIRO Agriculture, NSW, Armidale Australia
[2]NSW Department of Primary Industries, Beef Industry Centre, Armidale, NSW Australia

SUMMARY

Methane production in sheep is a novel trait that requires the development of consistent measurement protocols. The objective of this study was to estimate repeatabilities for methane production adjusted for liveweight measured in portable accumulation chambers in Merino ewes on pasture. Repeatabilities were low to moderate. No improvements in accuracy of the phenotypic variance could be achieved by additional measurements. Most likely the trait expressed at different ages and in particular different physiological status was not the same in lactating and dry animals, but the analysis in this study was not able to support this hypothesis.

INTRODUCTION

Methane emission from livestock could in future pose a constraint on freedom to operate if green house gas emissions are capped. As a novel target trait for ruminant livestock systems it calls for the development of measurement methods that are beyond current industry practice. Ideally that includes not just the measure of methane production but also of predisposing factors such as feed quality and intake leading to the amount of fermentable substrate. The objective of this study is to produce background knowledge for the development of a measurement protocol for methane production of ewes on pasture. Repeatabilities of methane production were estimated at different ages and physiological states and the increase in accuracy of measurement through repeated records investigated.

MATERIALS AND METHODS

Data on methane production was collected on 96 Merino ewes at different ages. The times of measurements reflect not only a trajectory in age, but the animals also differed in their physiological state (Table 1). Sheep were measured twice at approx. 15 months of age with 3-4 weeks between the two measurements (Treatment 1 and 2 (T1 and T2)), twice as lactating adults at about 21 months of age with 2 weeks between measurements (Treatment 3 and 4 (T3 and T4)) and once as dry adults at around 27 months of age (Treatment 5 = T5). Measurements were repeated once for each treatment within 3 days except for T5.

Table 1. Treatment names, age and physiological status of experimental sheep and pasture availability and time of year (Date)

Treatment	Age	Reproductive status	Repetition	Date	Pasture availability (kgDM/ha)
T1	Yearling (15m)	Dry	2	February 2013	1500
T2	Yearling (15m)	Dry	2	March 2013	1800
T3	Adult (21m)	Lactating	2	Early November 2013	915
T4	Adult (21m)	Lactating	2	Mid November 2013	1100
T5	Adult (27m),	Dry	1	May 2014	1100

Animals were kept on pasture at the Glen Innes Research and Advisory Station in the New England area of New South Wales, Australia. Feed availability varied from 900 to 1800 kg total dry matter per hectare (DM/ha) (Table 1). Methane production was measured using portable accumulation chambers (PAC), which enable individual animal measures of methane production in the field over a short period of time (Goopy *et al*. 2011). For this study individual animals were confined to the PAC for 40 or 60 minutes (Table 2). Liveweight (LWT) was recorded immediately after gas measurement. Animals were removed from feed and water one hour before measurement. Twelve sheep were measured per run, four runs were conducted each day. Animals were randomly assigned to runs and chambers in the order they entered the race. Therefore, short-term repeat measures after 3 days within each treatment, were recorded in a different order. It was not possible to record feed intake.

Statistical analysis. A univariate animal model for repeated measures was fitted using ASReml (Gilmour *et al.* 2009) to estimate repeatability (r) of methane emission in sheep. The repeatability is the ratio between the permanent environmental or between-animal variance (V_{Eg}) and the phenotypic variance (V_P), which is the sum of V_{Eg} and the temporary or within-animal variance (V_{Es}) (Falconer and Mackay 1996). It was not possible to fit a meaningful additive genetic effect with only four sires and limited pedigree.

Repeatabilities were estimated within and across treatments. Fixed effect levels within treatment comprised: day of measurement, run, repetition, chamber number and number of lambs at foot (none, single or twin lambs) for T3 and T4. Liveweight was fitted as a covariate to adjust for potential variation in feed intake and rumen volume. An identity matrix for the animal effect was fitted as random.

For the sake of comparison, we hypothesised that CH_4 in adult ewes (T5) is the most suitable measure to relate to lifetime CH_4 production, which was ultimately the trait that will become the breeding objective. The improvement in accuracy of phenotypic measurement is evaluated by adding measurements as lactating (T3 and T4) and young sheep (T1 and T2). Improvement was assessed by the associated effect on the phenotypic and environmental variances.

RESULTS AND DISCUSSION

CH_4 production is due to fermentation of feed in the rumen (Blaxter and Clapperton, 1965). As a consequence CH_4 production is expected to increase when more feed is ingested due to increased feed on offer or by increasing energy demand, e.g. lactation. Data was adjusted for liveweight, which was significant as covariate, but mean total CH_4 production differed significantly between treatments (Table 2). Total CH_4 production was highest during T3 and T4, despite low feed availability, because of increased feed intake due to the animals lactating during that time. Higher CH_4 production also occurred during T2 compared to T1 because more feed was on offer. Mean CH_4 production was the lowest for T5, most likely due to lowest intake as a consequence of amount of feed on offer and the ewes neither growing nor being pregnant or lactating.

Table 2. Descriptive statistics for methane production in Merino ewes at increasing age at pasture (mmol CH_4/min). Time = time period (mins) over which CH_4 was measured

Treatment	Time	No of records	Mean	Min	Max	StdDev
T1	60	192	0.91	0.38	1.59	0.23
T2	60	192	1.12	0.54	2.11	0.25
T3	40	192	1.32	0.55	2.47	0.38
T4	40	192	1.61	0.47	2.85	0.41
T5	40	96	0.84	0.39	1.30	0.22

Heritabilities (h^2) for methane production in sheep have been reported at h^2 ~0.29 for gCH4/day, 0.13 for gCH4/kg feed measured in respiration chambers (Pinares-Patino *et al.* 2013) and ~0.1 for gCH4/day adjusted for LWT measured in PACs (Robinson *et al*. 2014). Given the low to moderate heritabilities for methane related traits, the repeatability, which is the upper limit for the heritability, was expected to be moderate.

Repeatability with treatments. This study established that short term repeatabilities, measured within 3 days, for CH_4 adjusted for LWT were moderate for ewes at yearling age (r(T1)=0.33±0.09 and r(T2)=0.37±0.09). During lactation the repeatabilities were moderate (r(T4)=0.40±0.09) to high r(T3)=0.62±0.06. The increase in repeatability was due to higher between-animal and consequently phenotypic variance. In T5 all animals were only measured once and short term repeatabilities could not be established. The repeatability estimates were lower than repeatabilities from respiration chamber data on consecutive days (r = 0.94 ± 0.003, Pinares-Patino *et al.* 2013), which demonstrates the influence of controlled feed intake and highlights the problematic adjustment for LWT, as was done in this study.

Table 3. Repeatabilities and variances for CH_4 emission adjusted for LWT at different ages

Treatment	Repeatability	V_P	V_{Eg}	V_{Es}
T1 & T2	0.25 ± 0.07	0.027	0.007	0.020
T1 & T3	0.26 ± 0.06	0.039	0.010	0.029
T1 & T4	0.28 ± 0.05	0.047	0.013	0.034
T1 & T5	0.17 ± 0.05	0.029	0.005	0.024
T2 & T3	0.32 ± 0.08	0.034	0.011	0.023
T2 & T4	0.20 ± 0.05	0.043	0.009	0.034
T2 & T5	0.27 ± 0.06	0.026	0.007	0.019
T3 & T4	0.40 ± 0.07	0.057	0.023	0.034
T3 & T5	0.38 ± 0.06	0.037	0.014	0.023
T4 & T5	0.30 ± 0.06	0.047	0.014	0.033

Repeatability across/between treatments. Repeatabilities for CH_4 production adjusted for LWT across treatments, measured at least one month apart were low to moderate (Table 3). The estimates were lower than estimates reported by Pinares-Patino *et al.* (2013) of r=0.55±0.02 for gCH4/day, but align with estimates of r=0.25 for gCH4/day adjusted for LWT measured in PACs reported by Robinson *et al.* (2014). Repeatabilities are slightly higher at later ages, which was due to an increase in between-animal variance.

Low repeatabilities indicated that the accuracy of CH_4 measurement with PACs on animals from pasture would benefit from repeated measures. As outlined earlier, CH_4 emission at T5 was assumed to be the representative trait of life time CH_4 emission. It was investigated if the measures at different treatments were appropriate to add as repeated measures to increase the accuracy of the phenotypic variance. The results in Figure 1 demonstrate that any of the other treatments are unsuitable as repeated measures to increase the accuracy of phenotypic variance for CH_4 production in T5. CH_4 production in lactating ewes (T3 and T4) added variance, mainly through an increase in the within-animal variance. This could indicate that CH_4 production adjusted for LWT is a different trait in dry and lactating ewes. It also demonstrates that LWT might not be an appropriate adjustment for feed intake. This makes sense because lactating ewes would eat more and produce more CH_4 compared to dry ewes at the same LWT. A small decrease in phenotypic variance was observed by combining T5 and T2, but the addition of either T2 or T1 decreased the between-animal variance, which again, might be a reflection of a smaller additive genetic variance for T1 and T2 than T5. Differences in magnitude of the CH4 measurements between the

treatments would have contributed to the lower repeatability estimates when treatment data is added.

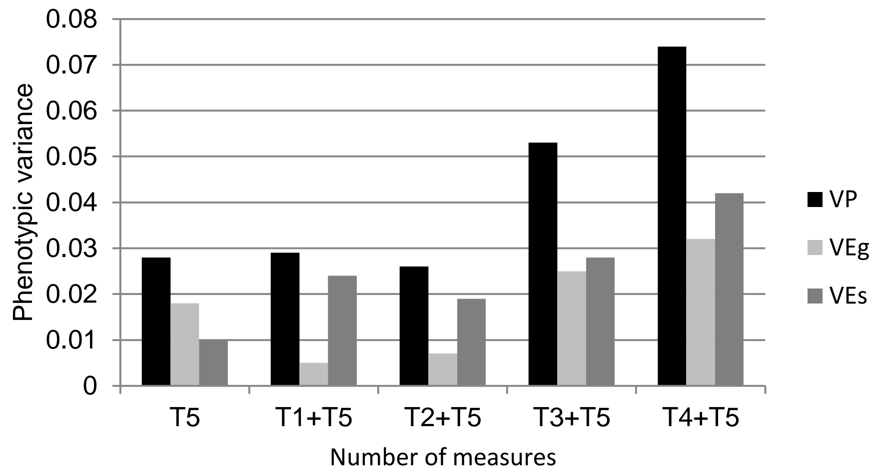


Figure 1. Phenotypic (V_p), permanent environmental (V_{Eg}) and special environmental variance (V_{Es}) with increasing number of measurements.

CONCLUSION

It is suggested that a measurement protocol for CH_4 production in Merino ewes on pasture in young or pregnant sheep is not a reliable indicator of adult performance. However, this data relates only to CH_4 adjusted for LWT and ignores the poor relationship between feed intake and LWT. A more desirable and appropriate phenotype for CH_4 production would account for the amount and quality of feed eaten, such as methane yield. However, it is not possible to measure feed intake with PACs in the field.

ACKNOWLEDGEMENTS

Without the hard work of A. Donaldson, M. Cameron, R. Woodgate, K. Quinn, P. Kamphorst and staff at NSW DPI Glen Innes Research and Advisory Station this work would not have been possible. Meat and Livestock Australia and Commonwealth Department of Agriculture contributed financial support for this work.

REFERENCES

Blaxter K.L. and Clapperton J.L. (1965) *Br. J. Nutr.* **19**: 511.

Falconer D.S. and Mackay T.F.C. (1996) ‘Introduction to Quantitative Genetics’ 4th Edition Longman Group Ltd.

Gilmour A.R., Gogel B.J, Cullis B.R. and Thompson R. (2009) ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempsted, UK.

Goopy J.P., Woodgate R.T., Donaldson A., Robinson D.L. and Hegarty R.S. (2011) *Anim. Feed Sci. and Tech.* **166-167**: 219.

Pinares-Patiño C.S., Hickey S.M., Young E.A., Dodds K.G., MacLean S. *et al.* (2013) *Animal* **7** (s2): 316.

Robinson D.L., Goopy J.P., Hegarty R.S., Oddy V.H., Thompson A.N. *et al.* . (2014) *J. Anim. Sci.* **92**:434

GENETIC PARAMETERS FOR METHANE PRODUCTION AND RELATIONSHIPS WITH PRODUCTION TRAITS IN AUSTRALIAN BEEF CATTLE

K.A. Donoghue[1], T.L. Bird-Gardiner[1], P.F. Arthur[2], R.M. Herd[3] and R.F. Hegarty[4]

[1] NSW Department of Primary Industries, Agricultural Research Centre, Trangie, NSW 2823
[2] NSW Department of Primary Industries, EMAI, Menangle, NSW 2568
[3] NSW Department of Primary Industries, Beef Industry Centre, Armidale, NSW 2351
[4] Environmental and Rural Science, University of New England, Armidale, NSW 2351

SUMMARY

This paper reports heritability estimates for methane traits and genetic relationships with production traits in beef cattle. Traits recorded during the methane test period included dry matter intake (DMI), test liveweight (TWT), methane production (MPR) and methane yield (MY; MPR/DMI). Two methods of calculating residual methane traits (RMP) were evaluated. Production traits included birth (BWT), weaning (WWT), yearling (YWT) and final (FWT) liveweight. Heritabilities for MPR, MY and RMP traits were moderate (0.19 to 0.27), indicating that there is potential to use genetic improvement to reduce methane emissions in Australian beef cattle. MPR was moderately genetically correlated with MY (0.50) and RMP traits (0.50 to 0.63). However, MPR was also moderately to highly genetically correlated (0.36 to 0.86) with weight traits. Methane yield and RMP traits, however, were lowly to moderately genetically correlated (-0.06 to 0.45) with weight traits. These results indicate that selection for lower MPR would have a negative impact on growth in beef cattle. Selection for reduced MY or RMP, however, would lead to reduced MPR with minimal impact on animal productivity. The use of a ratio trait, like MY, in animal breeding is generally undesirable, thus selection on RMP traits is a better alternative.

INTRODUCTION

Livestock make a significant contribution (14.5%) to greenhouse gas (GHG) emissions worldwide, and ruminants are the primary source (Gerber *et al.*, 2013). The use of genetic improvement to reduce GHG emissions would produce small, cumulative and permanent changes and would be particularly useful in extensive beef production systems. Methane production (MPR) has been found to be moderately heritable in sheep (Pinares-Patino *et al.*, 2011), and in preliminary reports from this study (Donoghue *et al.*, 2013; Herd *et al.*, 2014b). However, preliminary results indicate that MPR is highly genetically correlated to production traits (Herd *et al.*, 2014b). Alternative methane traits studied include methane yield (MPR/dry matter intake) and residual methane (difference between actual and predicted MPR), with both found to be moderately heritable (Donoghue *et al.*, 2013; Herd *et al.*, 2014b). The objective of this study was to quantify whether genetic variation existed for several methane traits, and to gain better understanding of the relationships between methane and production traits.

MATERIALS AND METHODS

Progeny born from Angus cows in 2 research herds at the New South Wales Department of Primary Industries Agricultural Research Centre, Trangie NSW, were measured for methane production in 10 respiration chambers on the University of New England campus, Armidale NSW. Herd *et al.* (2014a) provides details on the management of animals and methane measurement procedure. The 1,043 animals were progeny of 73 sires (average 14 progeny per sire, range 1-30), born across 4 drops. Each year, within herd and sex, cohorts of 40 head in 4 groups of 10 were formed, and progeny of individual sires were stratified across groups and cohorts.

Data. Methane production was measured over 2 x 24h consecutive periods. For animals born from 2011 to 2013 these measurements were taken at approximately yearling age (mean=339 days). However, for animals born in 2009, these measurements were taken at approximately two years of age (mean=738 days) due to delays in construction of the chamber facility. Traits measured included pre-test weight (TWT), dry matter intake (DMI), daily methane production (MPR) and methane production per unit feed intake (methane yield: MY). Two different forms of residual MPR (RMP) were defined to target MPR independent of feed intake, with RMP defined as actual MPR minus expected MPR (expMPR). For RMP_J, expMPR was calculated using a published prediction equation (Johnson *et al.*, 1995), while for RMP_R, the residuals from a simple regression of MPR on DMI were used.

Data for growth traits were collected on all animals in the research herds, including animals that had not been measured for methane. Growth traits recorded included birth (BWT), weaning (WWT), yearling (YWT) and final (FWT) weight, which were measured at birth and at mean (±SD) age of 231 (±23), 423 (±28), and 606 (±71) d, respectively. There were growth records available on 1,471 animals, who were the progeny of 75 sires (average 20 progeny per sire, range 1-38), though not all animals had all traits recorded. Editing of records included removal of animals with incomplete pedigree, missing birth date, large feed refusal during testing and trait measurements greater than 4 standard deviations from the contemporary group mean.

Model of analysis. Variance and covariance components were estimated with an animal model using ASReml (Gilmour *et al.* 2009). A fixed effect of contemporary group, random direct genetic effects, and residual effects were included in the standard model. Contemporary group was defined by cohort, methane group and management group. Covariates were added to the standard model where these variables were significant ($P<0.05$) for a particular trait. The standard model was used for RMP_R. For BWT, age of dam (in years) was added to the model as a linear covariate, while for FWT, a linear covariate for age of animal (in days) was included. For the remaining traits (TWT, DMI, MPR, MY, RMP_J,WWT and YWT), age of animal as well as age of dam were added to the model as linear covariates. For the traits of BWT, WWT and YWT, maternal genetic and maternal permanent environmental effects were also included in the model, with the direct-maternal genetic relationship fixed at zero. Pedigree records for all animals with records and 2 further generations of ancestors were used.

RESULTS AND DISCUSSION

Table 1 contains summary statistics for the methane test and weight data.

Table 1 Descriptive statistics for methane and growth traits

Trait	No. records	Average (SD)	Minimum	Maximum
TWT (kg)	1,043	356.4 (89.6)	156	640
DMI (kg/d)	1,043	6.07 (1.31)	3.59	9.42
MPR (g/d)	1,043	132.2 (25.4)	78.9	251.0
MY (g/kg DMI)	1,043	22.0 (2.3)	13.1	29.5
RMP_J (g/d)	1,043	10.7 (15.0)	-55.9	70.7
RMP_R (g/d)	1,043	0 (9.5)	-39.6	64.0
BWT (kg)	1,471	34 (4.8)	19	50
WWT (kg)	1,456	242 (37)	110	355
YWT (kg)	1,377	370 (54)	172	592
FWT (kg)	1,011	450 (58)	265	648

Genetic parameters for methane and production traits are reported in Table 2. Heritabilities for methane traits were moderate (0.27 and 0.22) and are similar to estimates of Donoghue *et al.*

(2013) and Herd *et al.* (2014b) using a smaller subset of the animals in this study. Pinares-Patino *et al.* (2011) also reported a moderate heritability (0.30) for MY in sheep. Heritabilities for RMP were moderate (0.19), similar to preliminary estimates from this study (Herd *et al.*, 2014b) and offer the potential to make selection decisions to target MPR independent of feed intake while also avoiding using a ratio trait, such as MY. The results from this study, together with published estimates, indicate that there is potential to lower methane emissions from livestock through selection.

Table 2 Genetic parameters (SE) for methane and growth traits

Trait	σ^2_d	σ^2_m	σ^2_c	σ^2_p	h^2_d	h^2_m	c^2
TWT	446.9 (98)	-	-	1,016 (54)	0.44 (0.08)	-	-
DMI	0.080 (0.017)	-	-	0.175 (0.009)	0.46 (0.08)	-	-
MPR	44.0 (12)	-	-	164.3 (8)	0.27 (0.07)	-	-
MY	0.383 (0.111)	-	-	1.76 (0.09)	0.22 (0.06)	-	-
RMP_J	15.7 (4.88)	-	-	84.1 (4.01)	0.19 (0.06)	-	-
RMP_R	15.7 (4.79)	-	-	83.8 (3.96)	0.19 (0.05)	-	-
BWT	6.32 (1.60)	3.31 (1.10)	0.32 (0.90)	18.38 (0.84)	0.34 (0.08)	0.18 (0.06)	0.02 (0.05)
WWT	172.6 (49)	73.7 (35)	95.7 (35)	670.8 (29)	0.26 (0.07)	0.11 (0.05)	0.14 (0.05)
YWT	465.7 (94)	48.6 (45)	30.8 (48)	1,002 (47)	0.46 (0.08)	0.05 (0.05)	0.03 (0.05)
FWT	827.5 (147)	-	-	1,390 (77)	0.60 (0.08)	-	-

Table 3 Genetic (above diagonal) and phenotypic (below diagonal) correlations (SE) for methane traits

Trait	TWT	DMI	MPR	MY	RMP_J	RMP_R	BWT	WWT	YWT	FWT
TWT		0.99 *(0.01)*	0.80 *(0.07)*	-0.10 *(0.18)*	0.05 *(0.19)*	-0.09 *(0.19)*	0.58 *(0.13)*	0.80 *(0.07)*	0.98 *(0.02)*	0.96 *(0.02)*
DMI	0.93 *(0.01)*		0.84 *(0.06)*	-0.04 *(0.18)*	0.10 *(0.18)*	-0.05 *(0.18)*	0.54 *(0.14)*	0.84 *(0.06)*	0.94 *(0.03)*	0.95 *(0.03)*
MPR	0.68 *(0.02)*	0.71 *(0.02)*		0.50 *(0.14)*	0.63 *(0.11)*	0.50 *(0.14)*	0.36 *(0.18)*	0.84 *(0.09)*	0.86 *(0.06)*	0.79 *(0.08)*
MY	0.04 *(0.04)*	-0.01 *(0.04)*	0.68 *(0.02)*		0.99 *(0.01)*	0.99 *(0.01)*	-0.01 *(0.21)*	0.27 *(0.21)*	0.21 *(0.18)*	0.05 *(0.17)*
RMP_J	0.11 *(0.03)*	0.08 *(0.04)*	0.76 *(0.01)*	0.97 *(0.01)*		0.99 *(0.01)*	0.03 *(0.22)*	0.45 *(0.20)*	0.38 *(0.17)*	0.18 *(0.17)*
RMP_R	0.02 *(0.04)*	-0.02 *(0.04)*	0.69 *(0.02)*	0.97 *(0.01)*	0.99 *(0.01)*		-0.06 *(0.22)*	0.32 *(0.22)*	0.23 *(0.19)*	0.06 *(0.17)*
BWT	0.43 *(0.04)*	0.39 *(0.04)*	0.26 *(0.04)*	-0.01 *(0.04)*	0.02 *(0.04)*	-0.03 *(0.04)*		0.53 *(0.15)*	0.56 *(0.12)*	0.54 *(0.14)*
WWT	0.71 *(0.03)*	0.71 *(0.03)*	0.53 *(0.03)*	0.03 *(0.04)*	0.11 *(0.04)*	0.04 *(0.04)*	0.36 *(0.04)*		0.92 *(0.04)*	0.92 *(0.05)*
YWT	0.85 *(0.01)*	0.80 *(0.02)*	0.61 *(0.03)*	0.09 *(0.04)*	0.16 *(0.04)*	0.08 *(0.04)*	0.40 *(0.04)*	0.66 *(0.04)*		0.99 *(0.01)*
FWT	0.84 *(0.01)*	0.79 *(0.01)*	0.56 *(0.03)*	0.10 *(0.04)*	0.13 *(0.04)*	0.07 *(0.04)*	0.39 *(0.04)*	0.62 *(0.03)*	0.84 *(0.01)*	

Phenotypic (r_p) and genetic (r_g) correlations and their associated standard errors between methane and growth traits are reported in Table 3. MPR was highly genetically correlated with both TWT (0.80) and DMI (0.84), indicating that reducing MPR would also lead to correlated reductions in TWT and DMI. In contrast, MY was not genetically correlated with TWT (-0.10) or DMI (-0.04), but was positively genetically correlated with MPR (0.50), indicating that reducing MY would have little impact on DMI or TWT, but have the correlated effect of reducing MPR. Large positive r_g (0.99) were observed between MY and the residual methane traits, indicating that, genetically, animals with higher MY also had higher RMP. Genetic relationships between the residual methane traits and TWT (-0.09, 0.05) and DMI (-0.05, 0.10) were low. This indicates that there is potential to select for reduced RMP with little impact on DMI and TWT, with the correlated effect of reducing MPR and the benefit of avoiding selection on a ratio trait. The genetic correlation between MPR and BWT (0.36) was moderate, while correlations with later growth traits (WWT, YWT and FWT) were large (0.79 to 0.86). These results are similar to those reported in sheep (Pinares-Patino *et al.*, 2013), where large genetic correlations were observed between MPR and WWT (0.71) and WT at 8 months of age (0.80). These correlations indicate that directly selecting for reduced MPR will also select for lighter animals. MY and residual methane traits were not genetically correlated with BWT or FWT (-0.06 to 0.18), but were lowly to moderately genetically correlated with WWT and YWT (0.21 to 0.45), however large standard errors were associated with all estimates. Pinares-Patino *et al.* (2013) reported little genetic relationship between MY and WWT (0.06) and WT at 8 months of age in sheep (0.06). The results in our study indicate that it may be possible to select for reduced MY or residual methane with minimal impact on animal productivity.

CONCLUSIONS

Genetic variation in methane emissions is present in this population of Angus cattle, confirming the potential to use genetic improvement to reduce methane emissions in livestock. For Australian beef cattle herds, selection for lower methane production (MPR) may lead to selection for lower weight and have detrimental effects on animal productivity. In contrast, selection for lower MY or RMP would lead to lower MPR with minimal impact on herd productivity. The use of a ratio trait, like MY, in animal breeding is generally undesirable, and thus selection on either of the residual methane traits is a better alternative.

ACKNOWLEDGMENTS

Project funds came from the Australian Government Department of Agriculture as part of its Carbon Farming Futures Filling the Research Gap Program, and Meat & Livestock Australia. We thank Simon Bird, David Mula, Glen Walker, Karen Dibley, Helen Smith, Chris Weber, Reg Woodgate and Kim Quinn for skilled technical assistance.

REFERENCES

Herd R.M., Arthur P.F., Bird S.H., *et al*. (2014a) *J. Anim Sci.* **92**:5267.

Herd R.M., Arthur P.F., Donoghue K.A., *et al*. (2014b) Proc. 10th WCGALP. Paper 038.

Gerber, P.J., Steinfeld, H., Henderson, B., *et al*. (2013). Tackling climate change through livestock – A global assessment of emissions and mitigation opportunities. FAO, Rome.

Gilmour A.R., Cullis B.R., Welham S.J. and Thompson R. (2009) 'ASReml User Guide Release 3.0'. VSN International Ltd, Hemel Hempstead, UK..

Johnson D. E., Hill T. M., Ward G. M., *et al.* (1995) Atmospheric methane: sources, sinks and role in global change. Springer-Verlag, New York, NY. p. 199.

Pinares-Patino C.S., McEwan J.C., Dodds K.G., *et al.* (2011) *Anim. Feed Sci. Tech.*. **166-67**:210.

Pinares-Patino C.S., Hickey S.M., Young E.A., *et al.* (2013) *Animal.* **7(s2)**:316.

GENOMIC ESTIMATED BREEDING VALUES FOR METHANE PRODUCTION IN AUSTRALIAN BEEF CATTLE

B.J. Hayes[1,2], K.A. Donoghue[3], C. Reich[1], B. Mason[1], R.M. Herd[4] and P.F. Arthur[5]

[1]AgriBio, Centre for AgriBioscience, Biosciences Research, DEDJTR, VIC 3083
[2]Biosciences Research Centre, La Trobe University, VIC 3083
[3] NSW Department of Primary Industries, Agricultural Research Centre, Trangie, NSW 2823
[4] NSW Department of Primary Industries, Beef Industry Centre, Armidale, NSW 2351
[5] NSW Department of Primary Industries, EMAI, Menangle, NSW 2568

SUMMARY

Methane emissions for beef cattle are heritable, whether measured as methane production, methane yield (methane production/dry matter intake), or residual methane (observed methane production – expected methane production). This suggests methane emissions could be reduced by selection. Genomic selection is perhaps the most feasible approach to implement for the beef industry, given the high cost of measuring methane production from individual cattle. Here we derive genomic estimated breeding values (GEBV) for methane traits from a reference set of 747 Angus animals measured for methane traits, and genotyped for 630K SNPs. The accuracy of GEBV was evaluated in a cohort of 273 Angus animals. Accuracies ranged from 0.29, for methane yield, to 0.35 for residual methane. Selection on GEBV using the genomic prediction equations derived here could reduce emissions for beef cattle by roughly 5% over 10 years.

INTRODUCTION

Methane emission levels, whether measured as methane production, methane yield (methane production/dry matter intake), or residual methane (difference between actual and predicted methane production) are all heritable traits (Donoghue *et al.* 2013; Herd *et al.* 2014). Selection for reduced emissions could therefore result in likely small annual but cumulative and permanent changes in emission levels. Residual methane production (RMP) or methane yield (MY) are more attractive targets for selection than methane production (MPR), as they are not unfavourably correlated with production traits (Donoghue *et al.* 2015).

Unfortunately given the cost and difficulty of measuring these traits, it is unlikely that either MY or RMP could be measured on the scale that would be necessary to calculate estimated breeding values (EBV) on an ongoing basis for the beef industry. An alternative is to use genomic selection for these traits. This entails measuring a large reference population for MY or RMP, genotyping the reference population for a large number of SNP markers, and then using the information to derive a genomic prediction equation, that can be used to calculate genomic estimated breeding values (GEBV) for any selection candidate that is genotyped. Here we use a large group of Angus animals measured for methane emission levels (as described by Donoghue *et al.* 2015), and real or imputed genotypes for 632,003 SNPs were used, to derive GEBV for MPR, MY and RMP. The accuracy of the GEBV was demonstrated to be moderate, enabling selection for reduced methane emission levels for Australian beef cattle.

MATERIALS AND METHODS

Phenotypes. For a full description of phenotypes, see Donoghue *et al.* (2015), in this proceedings. Briefly, 1,020 Angus animals were measured for methane production in 10 respiration chambers on the University of New England campus, Armidale NSW (Herd *et al.* 2014a) provides details on the management of animals and methane measurement procedure. The animals were progeny of 73 sires (average 14 progeny per sire, range 1-30), born across 4 drops.

Progeny of individual sires were stratified across groups and cohorts. Methane production was measured over 2 x 24h consecutive periods. For animals born from 2011 to 2013 these measurements were taken at approximately yearling age (mean = 339 days). However, for animals born in 2009, these measurements were taken at approximately two years of age (mean = 738 days). Traits measured (Table 1) included pre-test weight (TWT), dry matter intake (DMI), daily methane production (MPR) and methane production per unit feed intake (methane yield: MY). Four different forms of residual MPR (RMP) were defined to target MPR independent of feed intake, with RMP defined as actual MPR minus expected MPR (expMPR). For RMP_J, expMPR was calculated using a published prediction equation (Johnson *et al.*, 1995), while for RMP_R, the residuals from a simple regression of MPR on DMI were used.

Table 1. Definition of traits

Trait name	Abbrev-iation	Units	Definition
Test Weight	TWT	Kg	Pre-test weight
Dry matter intake	DMI	kg/day	Dry matter intake during methane measurement
Methane production rate	MPR	g/day	Methane produced
Methane intensity	MI	g/kg	MPR per unit TWT (MPR ÷ TWT)
Methane yield	MY	g/kg	MPR per unit DMI (MPR ÷ DMI)
Residual methane$_B$	RMP_B	g/day	MPR net of expected MPR (expMPR) from the DMI, with expMPR obtained by formula of Blaxter and Clapperton (1965)
Residual methane$_J$	RMP_J	g/day	MPR net of expected MPR from DMI, with expMPR obtained by formula of Johnson *et al*. (1995)
Residual methane$_I$	RMP_I	g/day	MPR net of expected MPR from DMI, with expMPR obtained by formula of IPCC (2006)
Residual methane$_R$	RMP_R	g/day	MPR net of expected MPR from the DMI, with expMPR obtained by regression of MPR on DMI

Genotypes. 1,020 Angus cattle, that have been measured for methane traits, were genotyped with either 777,000 SNPs Illumina Bovine HD Array (847 animals) or the Bovine 54,000 SNP50 array (173 animals). The SNP positions used were from bovine genome assembly UMD 3.1 (University of Maryland, College Park, MD). Stringent quality control procedures were applied to the data. Monomorphic SNPs and SNPs with less than 5 copies of the rare allele were removed. Then genotype calls with GenTrain score (GenCall) > 0.6 are high quality; below this value they were excluded. For the animals genotyped with the HD array, there were 650,934 SNPs genotyped at GenCall > 0.6. Furthermore, 343 mitochondrial SNPs, 1,124 Y chromosome SNPs, and 1,735 unmapped SNPs were excluded. SNPs with duplicate positions or dubious positions given linkage disequilibrium with surrounding SNPs were also removed. 632,003 SNPs remained. Samples (animals) were checked for excess heterozygosity (>0.4 is a sign of sample contamination), and had to have more than 90% of SNP with GenCall scores >0.6. All 1,020 samples passed these quality control criteria, and 97.9 % of SNPs were genotyped at GenCall > 0.6. Missing genotypes for animals genotyped with the 777K were imputed using Beagle3 (Browning and Browning 2009), and the same program was used to impute the animals genotyped for the 50K to 632,003 genotypes, after quality control on 50K genotypes as for the 777K genotypes

Genomic heritabilities and genomic breeding values. The models fitted to the data were as described by Donoghue *et al*. (2015), except that genomic relationships were used to describe relationships between animals:

$y = Xb + Zg + e$, where **y** is a vector of trait records (WT, DMI, CH4, MY, MI, RMP_B, RMP_J, RMP_I or RMP_R), **b** is a vector of fixed effects including contemporary group, age and dam age, **X** is a design matrix allocating records to fixed effects, **g** is a vector of genomic estimated breeding values (GEBV), **Z** is a design matrix allocating records to breeding values, and **e** is a vector of random residuals ~ N(0, $\mathbf{I}\sigma^2_e$), where σ^2_e is the error variance. The **g** were assumed distributed N(0, $\mathbf{G}\sigma^2_{gen}$), where σ^2_{gen} is the additive genetic variance and **G** is the genomic relationship matrix constructed from the 632,003 SNP markers genotypes, following Yang *et al.* (2010). Variance components were estimated on the full data set (1,020 records) using ASReml (Gilmour *et al.* 2009). Genomic heritabilities were then calculated as:

$$h^2 = \frac{\widehat{\sigma^2_{gen}}}{\widehat{\sigma^2_{gen}} + \widehat{\sigma^2_e}}.$$

The accuracy of genomic estimated breeding values (GEBV) was evaluated by predicting the youngest cohort of animals, those screened in 2014 (273). The reference population were then all the other animals (747). The accuracy of prediction was taken as for the animals in the validation set, the correlation of their genomic estimated breeding values and their phenotypes (corrected for fixed effects), divided by the pedigree heritability of the trait: $r(GEBV, y*)/\sqrt{h^2}$.

RESULTS AND DISCUSSION

The estimates of genomic heritabilities were very similar to those previously calculated using pedigree data (Donoghue *et al.* 2015) for most traits, and were within one standard error for all traits (Table 2).

Table 2. Estimates of heritability from analysis using either pedigree or genomic information to construct relationships between animals, and accuracy of genomic estimated breeding values in a validation cohort. Standard errors are in brackets.

Trait name	h^2 pedigree*	h^2 genomic	Proportion of genetic variance explained by SNP	Accuracy of GEBV
Weight (kg)	0.43 (0.08)	0.42 (0.07)	0.96	0.37
Dry matter intake	0.44 (0.08)	0.37 (0.07)	0.82	0.35
Methane Production	0.27 (0.06)	0.28 (0.06)	1.05	0.35
Methane Yield	0.22 (0.06)	0.20 (0.05)	0.92	0.29
Methane Intensity	0.28 (0.06)	0.25 (0.06)	0.83	0.29
Residual $methane_B$	0.19 (0.06)	0.18 (0.05)	0.97	0.30
Residual $methane_J$	0.19 (0.05)	0.18 (0.05)	0.98	0.34
Residual $methane_I$	0.19 (0.05)	0.18 (0.05)	0.96	0.34
Residual $methane_R$	0.19 (0.05)	0.18 (0.05)	0.94	0.35

*From Donoghue *et al.* (2015) using the same data.

The proportion of the additive genetic variance captured by the SNP (the estimated genetic variance from the SNP divided by the genetic variance estimated from pedigree) ranged from 0.82 to 1, and was close to 1 for most traits. This is encouraging, indicating the SNPs are picking up most of the genetic variation for the traits (the proportion of genetic variation explained by the SNP sets an upper limit on the accuracy of GEBV that can be achieved).

The accuracies of GEBV from GBLUP were moderate, and quite similar across traits (Table 2). Accuracies were all significantly different to zero - the standard error of the correlation between GEBV and phenotypes (which divided by square root of heritability gives the accuracy) was 0.06, and for all traits the correlation was positive and at least twice this standard error. The accuracies of GEBV are similar to those for methane traits in sheep (Rowe et al. 2014).

In conclusion, results were encouraging – accuracies of GEBV for all traits were moderate, even though no SNPs with large effects for any of the methane traits was observed. Given an accuracy of GEBV of 0.3 (e.g. for methane yield and methane intensity), we can calculate response to selection for these traits that could be achieved per year (very roughly) as:

$$\Delta G = \frac{ir\sigma_{gen}}{L}$$

where i is the intensity of selection (assume 1.5), L is the generation interval (assume 3.5), $r = 0.3$ is the accuracy of genomic breeding values, σ_{gen} is the genetic standard deviation for the trait. The selection response for methane yield and methane intensity would be 0.084 g/kg DMI and 0.002 g/kg live weight respectively. This is 0.4 % and 0.5 % of the mean for these traits – suggesting 10 years of selection could lead to a 4 % reduction in methane yield, or a 5 % reduction in methane intensity, using the genomic breeding values derived with the data set used here. This compares not too unfavourably with for example milk yield in dairy cattle, a much easier trait to measure, where roughly a 1.5 % gain per year is achieved.

REFERENCES

Blaxter, K. L., and Clapperton, J.L. (1965) *Brit. J. Nutr*. **19**: 511.

Browning, B. L. and Browning S.R. (2009) *Am. J. Hum. Genet*. **84**: 210.

Donoghue, K.A., Bird-Gardiner, T.L, Arthur, P.F, Herd, R.M., and Hegarty R.F (2015). *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21** (In press).

Gilmour A.R., Cullis B.R., Welham S.J. and Thompson R. (2009) VSN International Ltd, Hemel Hempstead, UK.

Herd R.M., Arthur P.F., Bird S.H., Donoghue K.A. and Hegarty R.S. (2014a) *J. Anim Sci.* **92**: 5267.

Herd R.M., Arthur P.F., Donoghue K.A., Bird S.H., Bird-Gardiner T. and Hegarty R.S. (2014b) *Proc. 10th WCGALP.* Paper 038.

Intergovernmental Panel on Climate Change (IPCC). (2006) Emissions from livestock and manure management. Chapter 10. In: S. Eggleston, L. Buendia, K. Miwa, T. Ngara, and K. Tanabe, editors, IPCC guidelines for national greenhouse gas inventories, volume 4: Agriculture, forestry and other land use. Institute for Global Environmental Strategies. Hayama, Japan. p.10.1-10.89.

Johnson D. E., Hill T. M., Ward G. M., Johnson K. A., Branine M. E., Carmean B. R. and Lowman D. W. (1995) Atmospheric methane: sources, sinks and role in global change. Springer-Verlag, New York, NY. p. 199.

Johnson, D. E., T. M. Hill, G. M. Ward, K. A. *et al.* (1995) Ruminants and other animals. In: M. A. K. Khalil, editor, Atmospheric methane: sources, sinks and role in global change. Springer-Verlag, New York, NY. p. 199-229.

Rowe, S., McEwan, J.C., Hickey, S. *et al.* (2014). *Proc. 10th WCGALP.* doi:10.13140/2.1.4751.2963

Yang, J., Benyamin, B., McEvoy, B.P. et al. (2010) *Nat. Genet*. **42**: 565.

DIVERGENT SELECTION FOR METHANE YIELD IN BEEF CATTLE

T. Bird-Gardiner[1], K.A. Donoghue[1], P.F. Arthur[2], R.M. Herd[3], and R.F. Hegarty[4]

[1] NSW Department of Primary Industries, Agricultural Research Centre, Trangie, NSW 2823
[2] NSW Department of Primary Industries, Agricultural Institute, Menangle, NSW 2568
[3] NSW Department of Primary Industries, Beef Industry Centre, Armidale, NSW 2351
[4] Environmental and Rural Science, University of New England, Armidale, NSW 2351

SUMMARY

Genetic selection and breeding to reduce methane production is one option to reduce greenhouse gas emissions, but correlated responses in production traits also need to be considered. The objective of this study was to quantify the effect of divergent selection for methane yield (MY), on methane and body weight traits in Angus cattle. High and Low MY selection lines were created in each of two performance-recorded Angus research herds during the 2011 mating season. This study is a preliminary report on the divergence of these selection lines, as assessed by the performance of the 2013 born progeny. There was no significant selection line by herd interaction. Approximately half a generation of selection was achieved. There was a significant ($P<0.05$) divergence between the two lines in the selected trait, methane yield. This was also reflected in the significant ($P<0.05$) selection line differences in the residual methane (actual minus expected methane production) traits and also in the estimated breeding values for these traits. There were no significant selection line differences in birth, weaning and yearling weights.

INTRODUCTION

Ruminants emit methane, a potent greenhouse gas (GHG). Methane is the main GHG emitted by ruminants and is a by-product of enteric microbial fermentation of plant material mainly in the rumen (McAllister et al. 1996). Hence reducing enteric methane production is essential to any GHG emissions reduction strategy in livestock. Higher feed intake is associated with higher methane production in ruminants (Blaxter and Clapperton 1965; Pelchen and Peters 1998). Feed intake is highly correlated with growth and other production traits in ruminants (Arthur *et al.* 2001; Lancaster *et al.* 2009). Therefore, breeding animals for lower methane production *per se,* may have a detrimental impact on ruminant productivity due to reduced feed intake. Consequently there has been increased interest in the amount of methane produced per unit feed intake, also known as methane yield (MY).

In 2009, a research project was started at the Agricultural Research Centre at Trangie, NSW, Australia, to investigate the potential of genetic improvement to reduce methane GHG emissions in cattle. Details of the project have been reported by Donoghue et al. (2015). This study is one of the components of the main project, and it provides a preliminary report on the performance of cattle divergently selected for methane yield at yearling age. The objective of this study was to establish if breeding could be used to reduce methane production and what effect this would have on methane and body weight traits

MATERIALS AND METHODS

Cattle utilized in this study were from a performance recorded registered Angus stud, comprised of two research herds (an Autumn calving and a Spring calving herd), located at the Trangie Agricultural Research Centre, NSW. As part of the main project, cattle born in 2009, 2011, 2012 and 2013 were measured for methane production in 10 respiration chambers at the University of New England campus, Armidale, NSW. For cattle born in 2009, males from both herds and females from one of the herds were measured for methane at approximately two years of

age (mean=748 days) in 2011 due to delays in construction of the methane test facility. For cattle born in subsequent years, animals from both sexes in both herds were measured as yearlings. Animals were on a restricted (alfalfa and oaten hay chaff) ration of approximately 1.2 times their estimated daily energy requirement for maintenance, when measured for methane. Details on the methane measurement protocols have been published earlier by Herd *et al.* (2014).

This study commenced in 2011 with the establishment of High and Low MY selection lines in each of the 2 research herds, comprising 330 females. Due to the limited number of females tested for methane in each of the two herds, untested females were also randomly allocated to the selection lines and included as the foundation animals. Therefore the 2008, 2009 and 2011 born animals formed the foundation herd for this study. Foundation females were randomly allocated to the High MY line (High MY) (174 cows) and the Low MY line (Low MY) (156 cows), irrespective of their individual MY values. All 16 bulls used for mating had methane measurements. Within herds, the four bulls with the highest MY were allocated to the High MY line and the four bulls with the lowest MY to the Low MY line. The sole selection criterion for bulls in the High MY line and Low MY line was individual MY. Throughout the project bulls and heifers were mated at approximately 14 months of age, and bulls were used for only one mating season. Animals from each selection line were grazed together throughout the year, except during mating. Allocation of mates within selection line was completely random, except for the avoidance of half-sib and son-dam matings. All matings were by natural service and the breeding herd were on pasture all year round, with supplementary feed offered during times of limited pasture growth. A total of 304 and 264 (dams and 2013 drop progeny) animals were tested for MY in the high MY and Low MY selection lines respectively. The progeny of the selected sires were born in 2013. Calves (287) were reared by their dams until weaning and were on pasture all year except during methane measurement. Pastures comprised native and introduced perennial and annual grasses and forbs (Windmill grass, *Chloris truncate*; spear grass, *Stipa spp*.; barley grass, *Hordeum leporinum*; burr-medic, *Medicago spp.*; and crowsfoot, *Erodium spp.*).

Traits studied. The definitions of all the traits studied are provided in Table 1. Methane production was measured over 2 consecutive 24 hour periods. Traits measured included pre-test

Table 1. Definition of traits

Trait name	Abbreviation	Units	Definition
Pre-test weight	TWT	kg	Weight at time of methane test
Dry matter intake	DMI	kg/day	Dry matter intake during methane test
Methane production	MP	g/day	Methane produced
Methane yield	MY	g/kg	MP per unit DMI (MP/ DMI)
Residual methane$_B$	RMP_B	g/day	MP net of expected MP (expMP) from the DMI, with expMP obtained by formula of Blaxter and Clapperton (1965)
Residual methane$_J$	RMP_J	g/day	MP net of expected MP from the DMI, with expMP obtained by formula of Johnson et al. (1995)
Residual methane$_I$	RMP_I	g/day	MP net of expected MP from the DMI, with expMP obtained by formula of IPCC (2006)
Residual methane$_R$	RMP_R	g/day	MP net of the expected MP from the DMI, with expMP obtained by regression of MP on DMI
EBV for MY	EBV_MY	g/day	Estimated Breeding Value for methane yield
EBV for RMP_R	EBV_RMP_R	g/day	Estimated Breeding Value for residual methane RMP_R
Birth weight	BWT	kg	Weight at birth
Weaning weight	WWT	kg	Weight at weaning (~ 200 d of age)
Yearling weight	YWT	kg	Weight at one year of age (~ 400 d of age)

weight (TWT), dry matter intake (DMI), daily methane production (MP) and MY. Body weights were taken at birth and approximately every 3 months, for research and routine husbandry practices. Estimated breeding values (EBV) for MY and RMP_R, generated for all animals in the main project (Donoghue *et al.* 2015), were also evaluated.

Statistical analysis. Data from the 2013-born progeny were analysed to study the responses to selection as they were the most advanced generation of selection. Analyses were conducted using ASReml (Gilmour *et al.* 2014), fitting generalized linear mixed models to evaluate fixed effects and sire fitted as a random effect. Fixed effects included herd, selection line and sex, with dam age and age at measurement fitted as covariates. Herd by selection line, herd by sex and selection line by sex were fitted as interactions. For the methane traits, test cohort (management test group within herd and sex) was also fitted as an additional fixed effect.

RESULTS AND DISCUSSION

There was no significant selection line by herd or selection line by sex interactions for any of the traits studied in the 2013 born cattle. This implies that the selection line responses were similar across the two herds and sexes. Methane test cohort was not significant for any of the methane traits. The 2013-born progeny were the first generation, but were only half a selection generation, as only one parent (sire) was selected on methane tests, dams were not allocated to selection lines on methane tests. Selection line means for the traits studied are presented in Table 2. There was a significant ($P<0.05$) divergence between the two selection lines in methane yield. There were also significant ($P<0.05$) selection line differences in residual methane traits and the EBVs for these traits. This difference is important given that it was achieved in half a generation of selection. This simulates what could be achieved at a commercial level, where only introduced sires/bulls are used to make genetic progress within the herd. There were no significant selection line differences in body weight.

Table 2. Least squares means (± standard errors) for methane production and growth traits of 2013 born cattle from the methane yield (MY) selection lines

Trait[1]	Selection line: High MY line	Selection line: Low MY line	Significance[2]
Number of animals	153	134	
Dry matter intake, kg/day	5.7 ± 0.1	5.7 ± 0.1	*ns*
Methane Production, g/day	127.4 ± 1.4	125.2 ± 1.5	*ns*
Methane yield, g/kg DMI	22.6 ± 0.1	22.1 ± 0.1	*
Residual $methane_B$, g/day	-13.5 ± 0.7	-16.7 ± 0.7	*
Residual $methane_J$, g/day	14.5 ± 0.7	11.4 ± 0.7	*
Residual $methane_I$, g/day	5.2 ± 0.7	2.1 ± 0.7	*
Residual $methane_R$, g/day	1.5 ± 0.7	-1.6 ± 0.7	*
EBV for MY, g/day	0.2 ± 0.1	-0.2 ± 0.1	*
EBV for RMP_R, g/day	0.8 ± 0.3	-1.4 ± 0.3	*
Birth weight, kg	31.3 ± 0.5	31.5 ± 0.5	*ns*
Weaning weight, kg	257.4 ± 3.8	259.9 ± 3.8	*ns*
Yearling weight, kg	416.2 ± 4.5	421.2 ± 4.5	*ns*

[1]See Table 1 for full trait names and definitions
[2] *ns* denotes non-significant difference ($P>0.05$); * denotes significant difference at $P<0.05$

Breeding cattle for lower methane production *per se,* may have a detrimental impact on productivity since low methane production is associated with a reduction in feed intake. Results

from this preliminary study support phenotypic and genetic correlation estimates between MY and live weight reported by Herd et al. (2014) and Pinares-Patino et al. (2013), through finding no significant difference between selection lines in live weight. Results of of this study reveal that selection for low methane yield (measured in respiration chambers, on restricted DMI) is possible and will result in a reduction in GHG emissions, which appear to have no impact on the growth of these Angus cattle. Further research is required to substantiate if these results are applicable to: unrestricted DMI of various feed types in cattle of various physiological state and breeds.

ACKNOWLEDGEMENTS

This work was funded by NSW Department of Primary Industries, University of New England, Meat and Livestock Australia and the Australian government, Department of Agriculture as part of the National Livestock Methane Program. Assistance provided by Simon Bird, David Mula, Glen Walker, Karen Dibley, Reg Woodgate, Helen Smith and Chris Weber is gratefully acknowledged. Statistical advice was graciously provided by Arthur Gilmour.

REFERENCES

Arthur, P.F., Archer., J.A., Johnston, D.J., Herd, R.M., Richardson, E.C. et al. (2001) *J. Anim. Sci.* **79:** 2805.

Blaxter, K.L., and Clapperton, J.L. (1965) *Brit. J. Nutr*. **19:** 511.

Donoghue, K.A., Bird-Gardiner, T., Arthur, P.F., Herd, R.M. and Hegarty, R.F. (2015) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21:** This proceedings.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. and Thompson, R. (2014) 'ASReml User Guide Release 4.0' VSN International Ltd, Hemel Hempstead, HP1 1ES, UK www.vsni.co.uk.

Herd, R.M., Arthur, P.F., Donoghue, K.A., Bird, S.H., Bird-Gardiner, T. et al. (2014) *J. Anim. Sci.* **92:** 5267.

Intergovernmental Panel on Climate Change (IPCC) (2006) IPCC Guidelines for National Greenhouse Gas Inventories, **4:** (*10*) 1. IGES, Japan.

Johnson, D.E., Hill, T.M., Ward, G.M., Johnson, K.A., Branine, M.E., et al. (1995) In 'Atmospheric methane: sources, sinks and role in global change',. pp. 199-299, editor M.A.K. Khalil, Springer-Verlag, New York.

Lancaster, P.A., Carstens, G.E., Crews Jr., D.H., Welsh Jr., T.H., Forbes, et al. (2009) *J. Anim. Sci*. **87:** 3887.

McAllister, T.A., Okine, E.K., Mathison, G.W. and Cheng, K.J. (1996) *Can. J. Anim. Sci*. **76**:231.

Pelchen, A. and Peters, K.J. (1998) *Small Rumin. Res*. **27:** 137.

Pinares-Patino, C.S., Hickey, S.M., Young, E.A., Dodds, K.G., MacLean, S., *et al*. (2013) *Animal* 7: (s2) 316.

BIG IS BEAUTIFUL: BIOLOGY INFORMED SEQUENCE EXPLOITATION

M. Pérez-Enciso[1,2,3], M. Naval-Sánchez[1], J. Leno-Colorado[2] and A. Reverter[1]

[1] CSIRO Agriculture Flagship, Queensland Bioscience Precinct, St Lucia, QLD 4067, Australia
[2] Centre for Research in Agricultural Genomics (CRAG), 08193 Bellaterra, Spain
[3] Institut Català de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

SUMMARY

We have entered the big data paradigm. Now that whole genome sequence data is available on a population scale basis, a fundamental issue is: what can be done with sequence data that cannot be achieved with former datasets? This question does not have a closed response, partly due to the fact that information contained in sequence data is highly repetitive (e.g., linkage disequilibrium) and also noisy (e.g., missing data due to shallow coverage). We argue that using accurate biology informed decisions can make a big difference in the prediction of genetic merit when sequence is available. Here we review the main kinds of external biological information and some approaches to combine these disparate sources. Despite the richness of resources available, two main difficulties lie ahead: (i) an improved understanding of the phenotype's biology to make the right the choice among the plethora of datasets available, and (ii) how this information is weighed and incorporated into selection decisions.

INTRODUCTION

The whole classical paradigm of animal breeding has been traditionally based on large datasets consisting of phenotypes and pedigree. Both kinds of information are rather homogenous and a unified, well accepted method was used for genetic evaluation, namely best linear unbiased prediction (BLUP). Molecular information in the form of low and high-density SNP arrays started to disrupt this data homogeneity. The amount of available molecular information in most livestock breeding programs has vastly increased recently, and this pace will only accelerate in the coming years. Today, the continuous decrease in sequencing and high performance computing (HPC) costs have made it conceivable the use of fully sequence in commercial breeding programs (Daetwyler *et al*. 2014).

Yet, it is important to realize that sequence data is not simply an increased SNP density. It is often said that, with sequencing, the causal mutations *are* in the data. But what is sometimes overlooked is that sequence data are very noisy, expensive to analyse, and error prone, especially at low coverage. As a result, derived genotypic data are highly unbalanced. For instance, in a large scale SNP discovery effort, where we analysed 120 pig genomes, only a few hundred SNPs out of all 45 million identified in total were called in all samples (Figure 1). This is to be the rule rather than the exception with this kind of data.

Despite initial enthusiasm based on simulation studies (Meuwissen & Goddard 2010), the limited empirical evidence on use of complete sequence for genomic selection so far calls for caution. Hayes *et al.* (2014) reported only a small (~4%) increase in accuracy compared to standard high-density array based selection. More recent simulations by Druet *et al.* (2014) and MacLeod, *et al.* (2014) suggest that the actual advantage will be heavily influenced by the allele distribution of causal variants and by recent demography (i.e. linkage disequilibrium). In parallel to the availability of larger genotype datasets and improved algorithms to predict genetic merit, vast amounts of new functional information are becoming available. After the sequencing of high quality reference genomes, gene expression datasets by RNA-seq across tissues are becoming available (e.g. Liao *et al*. 2014), and current and future essays on histone marks, methylation, open-chromatin transcription binding and chromatin conformation promise to unravel the

regulatory landscape governing biological processes (Andersson *et al*. 2015). An advantage of this kind of information is that it can, partly, be transferred across species (Villar *et al*. 2015). For instance, metabolic pathways are well conserved across mammals or even across eukaryotes for fundamental pathways as well as gene expression levels (Brawand *et al*. 2011). On the contrary, the regulatory levels across mammals are highly dynamic (Schmidt *et al*. 2010; Villar *et al*. 2015). Here, we review the different sources of current and foreseeable available information, and we suggest that careful utilization of this biological information might boost genomic selection.

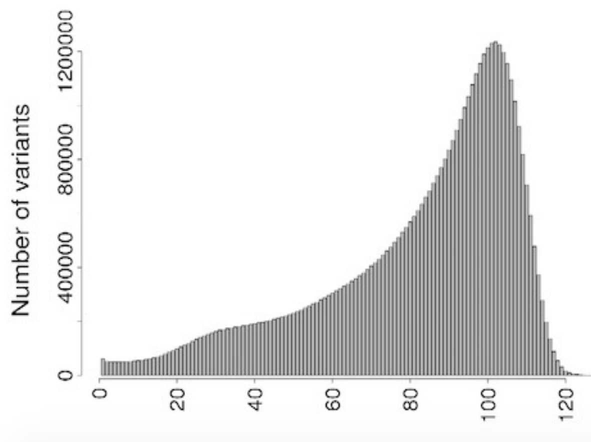


Figure 1: Number of individuals in which a given variant (SNPs) is observed. The data pertain to 128 pig genomes sequenced at varying depths, 4-20x, analysed with bwa and samtools. Figure from Bianco el al. (2015).

WHY FUNCTIONAL INFORMATION CAN BE USEFUL

Knowing the causal mutations is the holy grail for quantitative genetics. If these were known, much more accurate genetic predictions could be made, but note that this is but an extreme case of strong priors assigned to the SNPs available in the sequence dataset. In a recently published simulation study (Perez-Enciso *et al*. 2015), we showed that there is a clear law of diminishing returns when SNP density increases, and that the use of sequence would deliver only modest gains in accuracy. We predicted that only when using accurate biological information was sequence to pay off. Figure 2 shows our results. The two extremes are sequence data when used 'blindly', that is, without giving any different prior to any of the SNPs and inclusion of only the causal SNPs in the model. The latter strategy approaches an accuracy of 1, confirming our conjecture. Because all causal mutations are in the sequence, it is clear that wise choice of priors for each SNP can have a dramatic influence on prediction. Now, if all genes containing causal SNPs could be identified (red line) accuracy would increase by ~40%, as a result of disequilibrium with causal mutations. Yet, unfortunately, our simulations also show that miss- or incomplete specification of causal genes quickly diminishes accuracy (magenta and blue lines).

KINDS OF AVAILABLE INFORMATION

Table 1 shows a very shortlist of databases illustrating the wide diversity of data available that can be potentially used for improving the prediction of SNP functionality. These are: QTL, genome annotation, SIFT prediction, expression, methylation status, pathway information, gene ontology, among others. The new Functional Annotation of Animal Genomes (FAANG) consortium is currently gathering efforts to provide the same data to the animal genetics research community (www.faang.org) (Andersson *et al*. 2015) thus procuring a high quality genome annotation for domestic animals with unprecedented detail. Expression data are of particular

relevance. Defining which genes are expressed in which cell types and developmental time points is fundamental to our understanding of development and disease. RNA-seq data pictures whole genome expression levels independently of the species/breed of interest. Coupling differentially expression analysis together with motif discovery or pathway analysis results in further insight into regulation and biology. Seminal studies comparing the regulatory landscape across vertebrates have proven that regulatory regions are highly dynamic with only a core being conserved across species (Schmidt *et al*. 2010; Villar *et al*. 2015). Therefore, the regulatory annotation of distinct tissues and developmental stages across domestic species is crucial for their study.

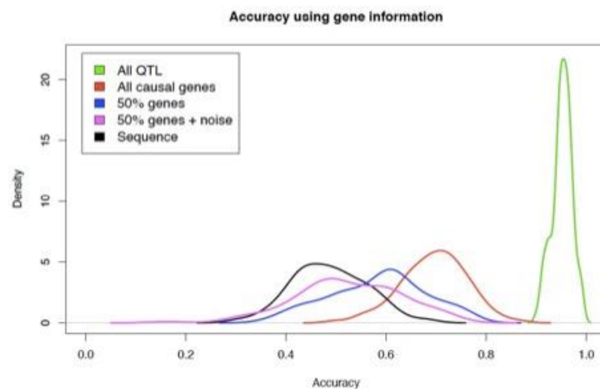


Figure 2. Comparison in accuracy between blind use of sequence (black) and if only QTL are employed in the model (green). The different curves show when either all SNPs within causal genes are used (red), 50% of the genes (blue) and added neutral SNPs (magenta). Details in Perez-Enciso et al 2015.

Related to this is the understanding of gene regulation itself. Motif discovery analysis on gene expression signatures is a popular alternative to detect regulatory regions and regulators in a given biological process. The underlying hypothesis to this strategy is that co-expressed genes tend to be co-regulated and therefore they might present similar Transcription Factor Binding Sites (TFBSs) in their regulatory region. Motif discovery analyses applied to differentially expressed (DE) gene sets predict changes in regulation. Differentially expressed genes are a consequence of a lack of functionality or mis-expression of a particular Transcription Factor, which triggers the downstream mis-expression of its direct target genes and a vast amount of indirectly related genes. Successful motif discovery tools in human are ModuleMiner, PhylCRM/Lever and i-Regulon (Janky *et al*. 2014; Van Loo *et al*. 2008; Warner *et al*. 2008).

In our opinion, the most promising approach for genomic selection is to utilize the information of how genes interact with each other, i.e., pathway analyses or 'gene set analyses' (GSEA). There are several tools for pathway analysis and, broadly, three kinds of functional pathway analysis: over-representation analysis, functional class scoring and pathway topology. Over-representation analysis requires that the input is a list of DE genes, this method evaluates the genes in a specific pathway that show changes in expression, counting the number of DE genes that are in the pathway. Functional class scoring analysis uses the entire data as input, this method follows three steps: first, computes differential expression of individual genes or proteins; second, the gene-level statistics of the genes of a specific pathway are aggregated into a single pathway-level statistic; finally, estimates the statistical significance of this pathway-level statistic. Pathway topology analysis uses the number and type of interactions between gene products, this method is essentially the same as the functional class scoring method but with the difference that the pathway topology analysis uses the additional information of the genes to compute the gene-level statistics.

COMBINING DISPARATE SOURCES OF INFORMATION

Formal integration of information from seemingly disparate sources aims at elucidating the congruency of these sources to further gain biological insight in a manner not possible by each individual source in isolation. The underlying premise is that inaccuracies are less likely to be present when separate data sources corroborate each other. Most applications for combining disparate sources in molecular biology follow one of three general approaches: meta-analysis, graph theory and cluster analysis. Here, we provide a brief description of each approach and when available present and discuss references of relevance to animal breeding and genetics.

Meta-analysis can be seen as an attempt to increase sample size. The objective is to achieve a higher statistical power by aggregating the results from separate studies linked by a common measure such as the effect of a SNP or the abundance of a gene. PRISMA is an organisation that provides guidelines for the systematic reporting of meta-analyses (http://www.prisma-statement.org/index.htm). Many journals endorse the PRISMA guidelines and require their authors to adhere to them. As an example, Pérez-Montarelo *et al.* (2012) undertook a meta-analysis of 20 gene expression studies in porcine spanning 134 experimental conditions on 27 distinct tissues. In an attempt to control the experimental design effects that may contribute to bias, the authors normalised the data by fitting a mixed-model approach that accounted for the disparity in the origin of the studies. With a focus on transcription factor genes and tissue-specific genes, a gene co-expression network was inferred where genes clustered by tissue and tissues clustered by embryonic origin. In another example, and in order to characterise inbreeding depression across species and traits, Leroy (2014) conducted a meta-analysis on 57 studies, 37 phenotypes and seven livestock species. Reported estimates of inbreeding depression were analysed using a multiple regression model that included the effect of study and phenotype. As result, the author reported an average decrease of 0.35% of the mean of a trait per 1% of inbreeding.

Graph-theoretic approaches have the intuitive appeal of network systems where objects (typically genes) are represented by nodes and relationships (typically interactions) are represented by edges. A number of attributes can be overlaid in the visualization schema and the resulting network visualized and explored using a (more or less friendly) software platform such as Cytoscape (www.cytoscape.org). Beyer and May (2003) developed a graph-theoretic algorithm, namely PARTITION, to the partition of individuals into full-sib families. Input to the algorithm is a list of individuals and their genotypes at each locus. For each pair of individuals, a likelihood ratio is calculated from the likelihood of being truly full-sibs over the likelihood that the pair is unrelated. The output is a list of full-sib families in the data set. A second example of graph-theoretic approaches is the work of Balasubramanian *et al.* (2004), who presented an approach for testing the association between multiple sources of functional genomics data, namely the edge permutation and node label permutation tests.

Finally, Bayesian correlated clustering (eg. Kirk *et al.* 2012), and Bayesian consensus clustering (Lock and Dunson 2013) are gaining momentum in the simultaneous integration of information from a wide range of different datasets and data types. In correlated clustering, the allocation of objects (e.g. genes) to clusters in one dataset has an influence on the allocation of genes to clusters in another dataset. Instead, consensus clustering is most commonly used to combine multiple clustering algorithms, or multiple realizations of the same clustering algorithm, on a single dataset.

TOWARDS A BIOLOGY INFORMED BREEDING ECOSYSTEM

The usefulness of sequence or high-density genotyping for genetic prediction is likely to reach a plateau rapidly, when used in isolation. In other words, there is so much redundancy in this kind of data that the likelihood ratio becomes flat when comparing alternative models with varying SNP density. In our opinion, the most promising way to move forward is by embracing the 'big data'

paradigm. However, contrary to what is normally understood by 'big data', the challenge is not in its size, but rather in its heterogeneity. Very much like internet companies try to make sense of the wide array of information collected by their clients in order to predict their behaviour, animal breeding companies should combine in an optimal way the huge public datasets containing biological information with their own phenotypic and polymorphism data. This is, admittedly, a vague recommendation and there is not, as of today, closed recipes to make the most of this information.

CAUTIONS

Even if very short and incomplete, this review points to the main issue that genomic selection will be facing if external biological information is to be successfully employed: how to weigh in an optimal way the vast diversity of external data that is already available. Our starting hypothesis is that there is not enough information in the data (i.e., in the likelihood) to tell whether a SNP is of sufficient relevance to be included or not in the predictive model and that, therefore, external information will be key to the successful use of sequence data.

REFERENCES

Andersson, L., Archibald, A.L., Bottema, C.D., *et al*. (2015) *Genome Biol*. **16**: 57.

Beyer, J., and May B. (2003) *Mol. Ecol*. **12**: 2243.

Balasubramanian, R., LaFramboise, T., Scholtens, D., *et al*. (2004) *Bioinforma. Oxf. Engl.* **20**: 3353.

Brawand, D., Soumillon, M., Necsulea, A., *et al*. (2011) *Nature* **478**: 343.

Daetwyler, H.D., Capitan, A., Pausch, H., *et al*. (2014) *Nat. Genet.* **46**: 858.

Druet, T., Macleod, I.M., Hayes, B.J. (2014) *Heredity* **112**: 39.

Hayes BJ, MacLeod IM, Daetwyler HD, *et al*. (2014). In *Proc.10th World Cong. Genet. App. Livest. Prod*.

Janky, R., Verfaillie, A., Imrichová, H., *et al*. (2014) *PLoS Comput. Biol*. **10**: e1003731.

Kirk, P., Griffin, J.E., Savage, R.S., *et al*. (2012) *Bioinforma. Oxf. Engl*. **28**: 3290.

Leroy, G. (2014) *Anim. Genet*. **45**: 618.

Liao, X., Bao, H., Meng, Y., *et al*. (2014) *PloS One* **9**: e102868.

Lock, E.F., Dunson, D.B. (2013) *Bioinformatics*. **29**: 2610.

MacLeod, I.M., Hayes, B.J., Goddard, M.E. (2014) *Genetics* **198**: 1671.

Meuwissen, T., Goddard, M. (2010) *Genetics* **185**: 623.

Perez-Enciso, M., Rincón, J., & Legarra, A. (2015). *Genet. Sel. Evol.*, in press

Pérez-Montarelo, D., Hudson, N.J., Fernández, A.I., *et al*. (2012) *PloS One* **7**: e46159.

Schmidt, D., Wilson, M.D., Ballester, B., *et al.* (2010) *Science* **328**: 1036.

Van Loo, P., Aerts, S., Thienpont, B., *et al*. (2008) *Genome Biol*. **9**: R66.

Villar, D., Berthelot, C., Aldridge, S., *et al*. (2015) *Cell*. **160**: 554.

Warner, J.B., Philippakis, A.A., Jaeger, S.A., *et al.* (2008) *Nat. Methods* 5, 347.

Table 1. Selected list of sites containing biological information

Database	Website	Description
Sequence		
GenBank	http://www.ncbi.nlm.nih.gov/entrez	An annotated collection of all publicly available DNA sequences.
EMBL	http://www.ebi.ac.uk/	Framework that provides free access to a range of mainstream sequence analysis applications.
DDBJ	http://www.ddbj.nig.ac.jp/	Primary nucleotide sequence database that provides analytical resources for biological information.
Protein		
SWISS-PROT	http://www.expasy.org/sprot	Swiss-Prot is the section of UniProtKB (central hub of protein knowledge) where the information is manually curated.
PIR	http://pir.georgetown.edu/	Resource that provides protein databases and analysis tools to support research on molecular evolution, functional genomics and computational biology.
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop	Database that provides a detailed and comprehensive description of the relationships of all known proteins structures.
Genomic		
Entrez biosystems	http://www.ncbi.nlm.nih.gov/biosystems/	Database providing integrated access to biological systems and their component genes, proteins, and small molecules, as well as literature describing those biosystems and other related data.
Entrez Genomes	http://www.ncbi.nlm.nih.gov/entrez	Database that contains sequence and map data from the whole genomes of over 1000 organisms.
KEGG	http://www.genome.ad.jp/kegg	Database of biological systems that integrates genomic, chemical and systemic functional information.
Organism-specific		
AnimalQTLdb	http://www.animalgenome.org/cgi-bin/QTLdb/	Contains reported QTL in livestock
FlyBase	http://flybase.bio.indiana.edu/	Database of genetic and genomic data for the insect family *Drosophilidae*.

OMIM	http://www.ncbi.nlm.nih.gov/Omim	Knowledgebase of human genes and phenotypes.
Transcription factor binding		
AnimalTFDB	http://www.bioguo.org/AnimalTFDB/	TF database specialized in livestock
TRANSFAC	http://transfac.gbf.de/	
DBD	http://www.transcriptionfactor.org/	Database of predicted transcription factors in completely sequenced genomes and their sequence specific DNA-binding domain families.
Epigenetic databases		
Epigenomics	http://www.ncbi.nlm.nih.gov/epigenomics	Resource for whole-genome epigenetic data sets.
MethDB	http://www.methdb.de/	Database for DNA methylation and environmental epigenetic effects.
The Histone Database	http://genome.nhgri.nih.gov/histones/	Resource for histones and histone fold-containing proteins.
CREMOFAC	http://www.jncasr.ac.in/cremofac/	Database dedicated for chromatin-remodeling factors.
Biochemical databases		
ENZYME	http://www.expasy.org/enzyme	Repository of information relative to the nomenclature of enzimes.
BRENDA	http://www.brenda-enzymes.org/	Database on functional and molecular information of enzymes.
AAindex	http://www.genome.ad.jp/dbget/aaindex.html	Database of phyisicochemical and biochemical properties of amino acids.

Proc. Assoc. Advmt. Breed. Genet. 21: 133-136

HIGH THROUGHPUT GENOTYPING-BY-SEQUENCING IN LIVESTOCK - ION AMPLISEQ^{TM}

Kristian Ridley[1], Jeremy Walker[2], Matthew Hickenbotham[3]
[1]Thermo Fisher Scientific, 33 Marsiling Industrial Estate Rd, Singapore
[2]GeneSeek, 48th St, Lincoln, NE, USA
[3]Thermo Fisher Scientific, 81 Wyman St, Waltham, MA, USA

SUMMARY

Genotyping-By-Sequencing (GBS) is rapidly gaining popularity for high throughput applications in livestock genetics and agriculture biotechnology. Ion Ampliseq^{TM} is a highly multiplexed, PCR-based resequencing technology that enables the targeting of hundreds to thousands of markers across hundreds of samples in a single sequencing run using the Ion Torrent^{TM} Sequencing. We recently demonstrated the power of Ion Ampliseq^{TM} GBS for high throughput cattle genotyping by designing a panel targeting over 4800 markers in a single pool. Furthermore, we have developed a simple, low cost, high throughput and rapid protocol ideal for commercial testing environments. The data shows excellent reproducibility, accuracy and >95% concordance with existing microarray data even at high (384) sample multiplex. GBS may also reveal novel variants within the targeted region that can enrich existing mapping data. Ion Ampliseq^{TM} offers the flexibility to design customized GBS panels for any species with a reference genome, and with little to no optimization required.

INTRODUCTION

Ion AmpliSeq^{TM} Designer is a simple assay design tool which can be used to create primer pools targeting any region of interest within a reference sequence. Amplicons ranging between 125 and 375 bp can be designed to target individual genetic loci such as single nucleotide polymorphisms (SNPs), or tiled across a total targeted region of more than 30,000 bp. For GBS applications, the reference genomes for mouse, cow, pig, sheep, dog, Chinese hamster, corn, rice, soybean, and tomato are preloaded into the Ion AmpliSeq^{TM} Designer. Additionally, Ion AmpliSeq^{TM} Designer allows the creation of Ion AmpliSeq^{TM} panels for private reference genomes or known target regions in a secure cloud computing environment.

With the simplicity and speed of PCR, Ion AmpliSeq^{TM} technology allows automated preparation of sequencing libraries. In combination with the Ion Proton^{TM} Sequencer, rapid molecular marker screening by GBS can be performed for hundreds of samples and thousands of targets in a single run. This study describes a collaborative investigation into the use of Ion Ampliseq^{TM} technology for developing a bovine SNP panel for a high throughput commercial testing application.

METHODS

Ion Ampliseq^{TM} panel design

Using the BosTau6 assembly with Y chromosome sequences from BosTau7 plus 2 custom contigs, a total of 4,874 SNP loci were submitted for to the Ion Ampliseq^{TM} Designer tool, of which 4,818 satisfied the design tool's *in silico* requirements (>99%).

Ion Torrent^{TM} Sequencing

Automated library preparation, including normalization using the Ion Library Equalizer^{TM} Kit was performed using the Tecan Freedom EVO® 150 platform using Ion Ampliseq^{TM} 2.0 reagents for 384

bovine gDNA samples using this custom primer pool. 192 or 384 libraries were then pooled prior to automated template preparation and chip loading using the Ion PI™ IC 200 Kit with the Ion Chef™ platform. Finally, samples were sequenced in duplicate using Ion PI™ Chip Kit v2 BC on the Ion Proton™ Sequencer.

Data analysis

Sequencing reads were aligned to the reference sequence using Torrent Suite™ Software under standard parameters and variants identified using the Torrent Variant Caller plug-in under default germline variant call parameters.

RESULTS

Call rates of 88% and 80% were observed at 192 and 384 sample multiplex respectively. The average genotype concordance to previous microarray data was >95% (r>0.95) for 4,469 positions in common between the two assays (excluding no-calls by both assays), with robust genotyping, regardless of sample multiplexing, of 192 or 384 libraries per sequencing run (Figure 1.). A set of 96 'beef diversity' samples, representing a wide range of cattle breeds was included to demonstrate the ability of GBS to identify polymorphisms within the amplicon target sequences surrounding the targeted SNPs. The numbers of novel SNPs identified by alignment of the sequencing reads to the bosTau6 genome assembly for all 384 samples is presented in Figure 2.

DISCUSSION

The study describes a novel genotyping-by-sequencing approach using Ion Ampliseq™ technology. This is the first account of multiplexed PCR targeted resequencing for a panel of SNPs of similar number to those currently used in low density bovine arrays (3000-7000 markers) for applications such as genomic selection. The high level of concordance and sample multiplexing demonstrated here indicate that Ion Ampliseq™ may represent an ideal solution for developing both smaller and larger bovine SNP panels for a range of livestock applications including parentage, inherited disease and key production trait testing. Ion Ampliseq™ panels can easily be modified by addition or subtraction of primers 'on-the-fly' without the need for extensive re-optimization, and this represents a major advantage over existing array-based or mass spec-based genotyping platforms. Furthermore, genotyping-by-sequencing generates additional SNP data from the amplicon fragments surrounding the targeted SNP, providing researchers with additional variant data that may aid mapping studies.

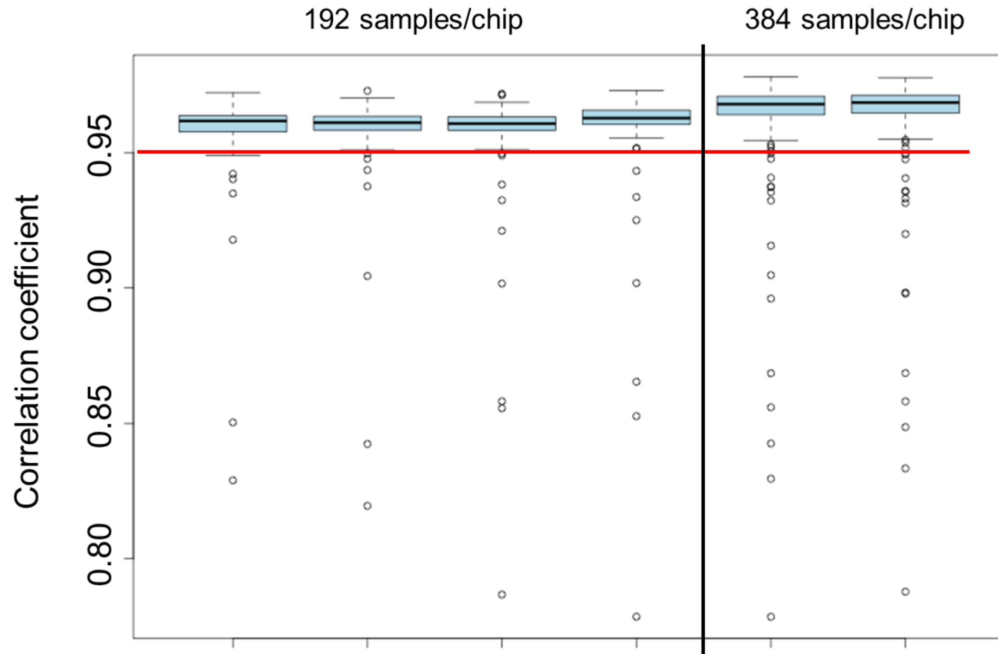


Figure 1. Genotype concordance data between an Ion AmpliSeq™ GBS panel and bead-based microarray analysis.

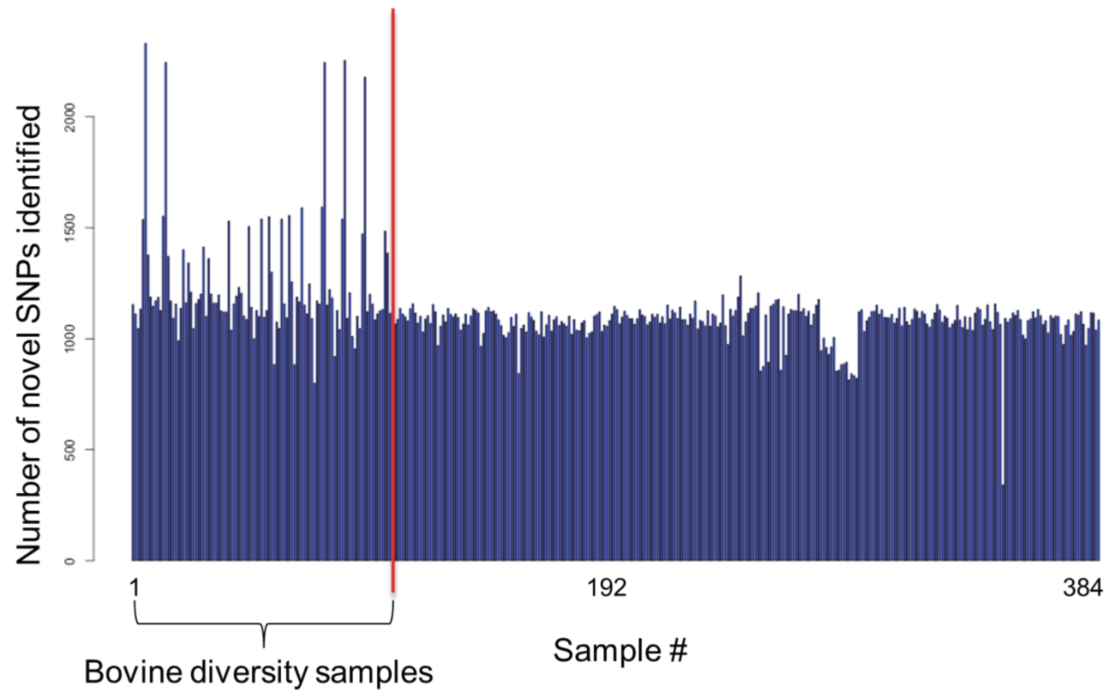


Figure 2. Number of novel SNPs identified by Ion Ampliseq™ sequencing for all 384 samples tested, including a set of 96 bovine diversity samples.

ACKNOWLEDGEMENTS

We would like to thank GeneSeek (Neogen Corporation) for providing reference sequences, target sequences and samples, and for their contribution to data analysis. The author would also like to acknowledge the staff involved in this study from Thermo Fisher Scientific (US) for design of the Ion Ampliseq™ panel and for conducting the Ion Torrent™ sequencing and analysis.

Proc. Assoc. Advmt. Breed. Genet. 21: 137-140

USING GENOMIC INFORMATION TO ESTIMATE GENOTYPE BY ENVIRONMENT INTERACTIONS

S.A. Clark[1,2] , S.H. Lee[1,2] and J.H.J. van der Werf[1,2]

[1] School of Environmental and Rural Science, University of New England, Armidale, NSW 2351
[2] CRC for Sheep Industry Innovation, University of New England, Armidale, NSW 2351

SUMMARY

Genotype by environment (GxE) interaction can reduce genetic gain because there is often insufficient information for accurate selection in each environment. Traditionally the estimation of GxE effects has been based on the performance of half siblings across environments. This limits the estimation of GxE to specifically designed datasets with close relatives where all realized relationships may not be utilized. Genomic information can also be used to link animals and presents an opportunity to compare genotypes across different environments using realized relationship information. This study examines the use of genomic information to estimate GxE interaction. The genetic correlation between animal phenotypic performance in two different environments was estimated using pedigree or genomic information. A higher genetic correlation between environments was observed when using genomic information (0.9) than when using pedigree information (0.71). This study suggests that genomic information may be a useful alternative to pedigree information in understanding GxE in livestock populations.

INTRODUCTION

In livestock production, animals are recorded and selected in a wide range of environments. While for most economically important traits there is little evidence for genotype by environment interaction (GxE), for some traits, animals or genotypes may perform differently in each environment (i.e. across different geographic locations or from one year to another). This can involve a change in the differences between alternative genotypes (often referred to as scale effects) and it can also relate to a change in the ranking of genotypes across environments.

Traditionally, GxE interactions can be estimated by measuring relatives across environments. Genotype by environment interactions can be estimated using mixed model analyses treating performance across environments as two different traits (Falconer 1952) and estimating the genetic correlation between performances across each environment. Past studies examining GxE have been limited to experimental designs that primarily focus on the use of common sires across various environments. The advent of genome-based technologies allows for the possibility of changing the way GxE interactions may be estimated.

Genome-wide association studies and genomic prediction have become common place for the prediction of disease risk in human populations and for predicting genetic merit in livestock (Goddard 2012). Genome-wide association and genomic prediction rely on a group of individuals with both genotypic and phenotypic information to enable the prediction of marker effects (directly or indirectly). Often these phenotypes come for a wide range of environments and the genomic information can be used to define the covariance between relatives (in the form of a genomic relationship matrix (GRM)) (VanRaden 2008). Genomic information presents an opportunity to enable a more diverse range of animals, not just close pedigree relatives, to contribute to estimating a genetic correlation between environments. It also presents the opportunity to observe whether specific genomic regions are more important than others for performance in varying environments.

The aim of this study was to use genomic information to estimate GxE and to examine the impact of such information when compared to pedigree based estimates.

METHODS

The data used in this study consisted of phenotypic and genotypic records from Merino animals in the Australian Sheep Cooperative Research Centre (CRC) information nucleus flock (INF). The INF is a specifically designed dataset that includes animals that have been recorded in eight environments across Australia. This dataset consisted of a dataset of phenotypic and genotypic records from 4433 Merino animals for the Post Weaning Weight (PWWT) trait. This dataset was further broken down such that phenotypic data from 1807 animals from 227 sires measured across two environments were extracted for the analysis. Location 1 (E1), Armidale (NSW) is a temperate environment with a primarily summer-dominant rainfall (n=921) and Location 2 (E2) Katanning (WA) is located in a winter-dominate rainfall zone (n=886).

All animals in each dataset were genotyped using the Illumina 50K ovine SNP chip. All SNP in this dataset underwent a number of genotyping quality control measures (see Daetwyler et al. (2010) and 48 599 markers remained following the quality control.

Genotype by environment interaction was estimated using both pedigree and genomic information. Phenotypic performance in the two environments was modelled as two separate traits. A bivariate animal model was fitted in ASReml (Gilmore 2009) and the genetic correlation between performance across environments was estimated using either a genomic or pedigree based relationship matrix. The following fixed effects were fitted in the analysis of PWWT: Sex, birth type, rearing type, age of dam, contemporary group (birth year • site • management group) and age-at-trait recording. We assumed the following model;

$$y_i = X_i b_i + Z_i a_i + Q_i s + e_i \ (1)$$

where $\mathbf{y}_i$ is a vector of phenotypes for environment i, $\mathbf{X_i}$ is a design matrix relating the fixed effects (as described above) to each animal for environment i, $\mathbf{b}_i$ is a vector of fixed effects, $\mathbf{Z}_i$ is a design matrix allocating records to breeding values, **a** is a vector additive genetic effects for animals, $\mathbf{Q}_i$ is a matrix relating animals to genetic groups and **s** is a vector of genetic group effects and $\mathbf{e_i}$ is a 2x2 diagonal matrix of random normal deviates $\mathbf{I}\sigma^2_{ei}$. Furthermore V(a) = $\begin{bmatrix} A\sigma^2_{a1} & A\sigma_a \\ A\sigma_a & A\sigma^2_{a2} \end{bmatrix}$ where σ^2_{ai} if the genetic variance for environment i and σ_a is the covariance between environments and A is the numerator relationship matrix. In the genomic analysis, the genomic relationship matrix (GRM) replaced the A such that V(g) = $\begin{bmatrix} G\sigma^2_{g1} & G\sigma_g \\ G\sigma_g & G\sigma^2_{g2} \end{bmatrix}$ (VanRaden 2008).

Marker effects for each environment were also estimated using single marker regression using the R package lm. The model fitted was

$$y = Xb + Q + SNP_j + e_i \ (2)$$

Again **y** is a vector of phenotypes, **X** is a design matrix relating the fixed effects (as described above) to each animal. Genetic groups (**Q**) were fitted as fixed effects. Each SNP was individually fitted until all markers had been tested. In this analysis three groups of phenotypic data were used to estimate the marker effects; the complete INF dataset of Merino animals with PWWT records (n=4433) across all eight environments, records from E1 (n=921) and records from E2 (n=886). The 500 most significant markers from E1 and E2 were then used to estimate a correlation across environments.

RESULTS

A moderate genetic correlation between environments was estimated for PWWT in Merino sheep using pedigree information (Table 1). By contrast, when the GRM was used to define the

covariance between individuals the genetic correlation between performances across environment was higher. Similar variance components were estimated for E2 when using either genomic or pedigree information. However, large differences between the variance component estimates were observed for E1 which contributes to the variable genetic correlation estimates. There was a high standard error surrounding each genetic correlation such that each correlation was not significantly different however it is interesting to observe such large dissimilarities between the estimates.

Table 1. Genetic variance (V_a), Phenotypic variance (V_p), heritability (h^2) and genetic correlation (r_g) of performance across alternative environments using either pedigree or genomic information (S.E).

	Pedigree				**Genomic**			
	E1	**E2**	**cov**	**r_g**	**E1**	**E2**	**cov**	**r_g**
V_a	10.35	13.00	8.32	0.71 (0.18)	6.79	12.15	8.192	0.90 (0.15)
V_p	15.55	23.04			15.19	22.77		
h^2	0.66 (0.11)	0.56 (0.12)			0.44 (0.074)	0.53 (0.085)		
LogL	-3458.36				-3434.29			

The marker effects were different across environments (Figure 1). When information from eight environments was used to estimate marker effects a large significant peak was observed on Chromosome 6. This location is consistent with that reported by Al-Mamun et al (2014). The strength of this peak reduced when the dataset was limited to either location 1 or 2 information (Figure 1b and 1c). Figure 1d shows the relationship between the effects of the most significant markers from each environment (E1 and E2).

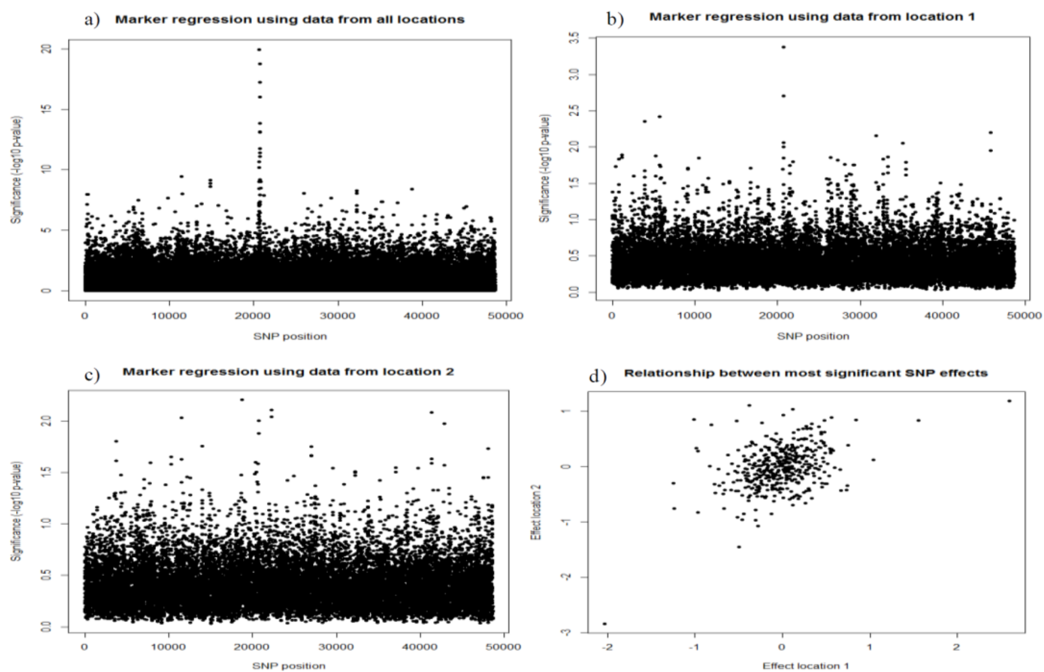


Figure 1 Manhattan plot of marker significance using data a) from all eight INF environments b) from location 1 c) from location 2. d) The relationship between the 500 most significant SNP from b & c.

The correlation between the SNP effects of significant markers estimated from each environment was lower than that estimated from the bivariate analysis (0.39). Estimating marker effects from individual environments is somewhat problematic given the reduction in the number of records available to detect marker effects, which can only be improved by increasing the amount of data used to estimate such effects. Combining data from many environments allowed for greater statistical power to be achieved and for a significant region to be observed. If a significant GxE interaction was to exist, marker effects may also be affected by this interaction and therefore may not result in consistent predictions across environments (if data was separated into specific environments). Furthermore, the estimated correlation between significant effects may not be a true reflection of the actual genetic correlation across environments due to the high degree of similarity between significant markers (i.e. many markers are in fact tracing the same genomic region). There would have also been a large amount of Linkage disequilibrium between markers due to the structure of the data. This could be corrected for by fitting all markers within the model (i.e. RR BLUP) but given the equivalence between gBLUP and RRBLUP (Habier et al 2013) the current gBLUP analysis would have resulted in a better estimate of the genetic correlation between environments.

The reasons for the disparity between pedigree and genomic estimates are not completely clear. The Log likelihood from each analysis suggests that using genomic information was in fact a better model, significantly increasing the likelihood of the data. This increase, however, may have been due to a number of factors. The first explanation is that the GRM may have better parameterized the relationship between the commercial dams that were used to create this dataset and better corrected for the genotypic effects across environments than what was captured by pedigree. A second explanation is that the GRM may have also included some genetic group information that was not available to the pedigree based matrix and could not be separated from the GRM. This would imply that the genomic estimate may be overestimating the genetic correlation across environments.

CONCLUSION

Genotype by environment interactions can be estimated using either pedigree or genomic information. Genomic information allows for the comparison of all animals across environments, not just animals from sire families or that have close pedigree links. Estimates of GxE may be different when comparing pedigree or genomic relationships and careful consideration needs to be made when interpreting such differences.

REFERENCES

Falconer, D.S. (1952). *Amer. Nat*. **86**:293–298

Gilmour, A.R., Gogel, B.J., Cullis, B.R. *et al*. (2002). ASReml User Guide Release 3.0 VSN International.

VanRaden, P.M. (2008). *J. Dairy Sci.* **91**: 4414.

Daetwyler, H.D., Hickey, J.M., Henshall, J.M., Dominik, S., Gredler, B. *et al*. (2010). *Anim. Prod. Sci.* **50**: 1004.

Al-Mamun, H.A., Clark, S.A., Kwan, P., Gondro, C. (2014) Proc. 10th WCGALP.

Goddard, M.E. (2012) *Anim. Prod. Sci.* **52**: 73.

Habier D., Fernando R.L., Garrick D.J., (2013) *Genetics* **194**(3): 597.

Proc. Assoc. Advmt. Breed. Genet. 21: 141-144

LAMB EATING QUALITY CAN BE MANAGED BY USING TERMINAL SIRES WITH DESIRABLE BREEDING VALUES FOR INTRAMUSCULAR FAT AND SHEAR FORCE

J.E. Hocking Edwards[1,4], B. Hancock[2,4], S. Gill[3,4], R. Apps[3,4] and A. Ball[3,4]

[1]South Australian Research and Development Institute, PIRSA, Struan Research Centre, Naracoorte, SA, 5271
[2]Rural Solutions SA, PIRSA, Roseworthy Campus, Roseworthy, SA, 5371
[3]Meat & Livestock Australia, Armidale, NSW, 2351
[4]Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW 2351, Australia

SUMMARY

The lamb industry has made significant genetic change in growth rate, leanness and muscling, but has not been able to efficiently effect genetic change in eating quality. Research breeding values (RBVs) for intramuscular fat (IMF) and shear force (SF5) have been developed for the Australian sheep industry and the effectiveness of these was determined in 16 prime lamb production systems. Ewes were inseminated with semen from rams with divergent RBVs for IMF and SF5 and their lambs were processed through 13 abattoirs for seven lamb supply chains.
A 1% increase in terminal sire IMF RBV resulted in a 0.57% increase in lamb IMF and a 1N decrease in sire SF5 RBV resulted in a 0.7N decrease in shear force. However, there were unfavourable effects of these eating quality RBVs on other carcase traits; in particular, selection for improved eating quality using SF5 or IMF RBV is likely to decrease lean meat yield (LMY). Therefore, both eating quality and other carcase traits need to be taken into consideration simultaneously in genetic improvement programs in terminal lamb productions systems.

INTRODUCTION

Eating quality of lamb meat is largely driven by tenderness, juiciness and flavour, with consumers both domestically and internationally demanding premium quality and value for money when purchasing prime lamb meat (Pethick *et al.*, 2011). Meat tenderness, measured objectively as shear force at five days aging (SF5) and intramuscular fat (IMF) are the two key traits that determine eating quality and therefore consumer satisfaction for lamb (Pannier *et al.*, 2014) and both traits are heritable (IMF h^2=0.48; SF5 h^2=0.27; Mortimer *et al.*, 2014). The lamb industry has made significant genetic change in growth rate, leanness and muscling, but has not been able to efficiently effect genetic change in eating quality (EQ) due to the difficulty in measuring and selecting for these traits. In addition, there is a growing concern that the use of sires that are superior for lean growth might have a negative impact on the eating quality of lamb meat for consumers. Genetic selection for leanness and muscling has been linked to declining IMF levels (Hopkins *et al.*, 2005), which can have detrimental effects on the eating quality of lamb (Pannier *et al.*, 2014). The development of research breeding values (RBVs) for hard to measure traits by the Sheep CRC and Sheep Genetics may enable the sheep industry to genetically manage eating quality of lamb (Daetwyler *et al.*, 2012). Meat quality traits lend themselves particularly well to genomic prediction given they are currently impossible to measure in a live sheep. The aim of this project was to deliver “proof of concept” for eating quality attributes within major lamb supply chains and to determine the impact that selection of sires for these newly generated RBVs will have on eating quality of their lamb progeny in a commercial production system.

MATERIALS AND METHODS

Producer demonstration sites (PDS) were located in Western Australia (N=2), South Australia (N=2), Victoria (N=6), Tasmania (N=3) and New South Wales (N=3). Animal use in the project was approved by the respective organisational Animal Ethics Committees.

Commercial lamb producers prepared ewes for an artificial insemination (AI) program and commercial AI operators were engaged to undertake the process. Composite, Merino, White Suffolk x Merino, Corriedale, Cormo and Coopworth ewes (N=5752) were mated with semen from terminal sires (Poll Dorset and White Suffolk). Rams were selected for divergent RBVs for IMF and SF5. RBVs were calculated using single step genomic prediction that included all known genomic information from sheep with a 50K SNP test and all phenotypic information collected from the Sheep CRC Information Nucleus and Resource Flocks. The IMF RBVs ranged from -0.89% to 1.21% between sires and the SF5 RBVs ranged from -5.3N to 6.4N. Eight rams were used at each site with the exception of one site where semen from one sire was unviable so only seven rams were used at this site. A total of 86 terminal sires (39 Poll Dorset, 47 White Suffolk) and one maternal sire (Corriedale) were used, with 24 terminal sires (9 PD; 15WS) used at more than one site. Sires were given equal opportunity within site with ewes randomised for weight and body condition score. Sire RBVs were provided by Sheep Genetics from a run completed in September 2014, which did not contain data from the progeny in this experiment. A small blood sample was collected from the ear of each lamb at marking and sent to a commercial provider for parentage testing (sire only). Lambs were finished under normal commercial conditions to meet the individual producers target market and were slaughtered at 13 different plants. Carcase and eating quality measurements were undertaken in accordance with those developed by the Sheep CRC (Pearce, 2009). The carcases had an average IMF of 4.05% (SD=0.852%, min=1.42%; max=7.98%) across the 1303 lambs measured and PDS averages ranged from 3.33 ± 0.893% to 4.81 ± 0.865% across the 16 PDS. The average shear force was 36.9N (SD=13.63N; min=14.2N; max=100.1N) across the 1292 lambs and PDS averages ranged from 23.9 ± 7.89N to 53.4 ± 13.66N across the 16 PDS.

Statistical Analysis. IMF and SF5 data were analysed with a linear mixed effects model (SAS v9.3, SAS Institute, Cary, NS, USA). The model included site, kill group within site, birth type within site (single, multiple, unknown), sex and breed (PD, WS) as fixed effects. The curve linear term for each RBV along with interactions with sex and farm were also included in the models. Sire was included as a random effect. HCWT and its interaction were included as a covariate. The sire solutions from the analysis of each trait were estimated from the model with the RBV removed.

RESULTS & DISCUSSION

Relationship between IMF RBV and progeny performance. Sire was a significant covariate for IMF (P=0.0002). When IMF RBV was included as a covariate, the RBV had a significant effect on progeny IMF (P<0.0001). Across a 1.5% IMF RBV range, progeny IMF increased by 0.86 units of IMF, resulting in a 0.57 ± 0.097% increase in IMF associated with 1% increase in IMF RBV (Figure 1a). The use of the IMF RBV is likely to illicit a more rapid change in IMF levels than using PFAT ASBV which achieves between 0.1% IMF to 0.17%IMF per mm PFAT (Pannier *et al.*, 2014; Hopkins *et al.*, 2007).

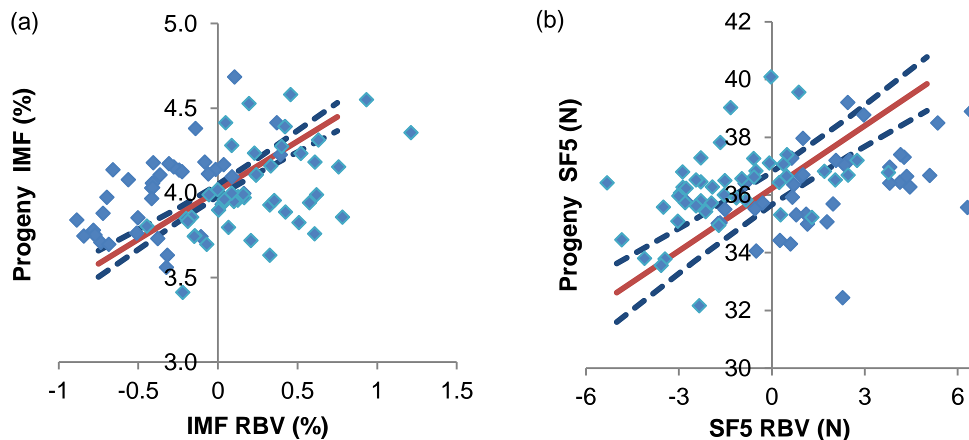


Figure 1. Relationship between (a) intramuscular fat (IMF) RBV and sire estimate of progeny IMF and (b) shear force (SF5) RBV and sire estimate of progeny SF5. Solid lines represent least square means of the sires RBV and dashed lines are the SEM. Sire estimates are obtained from the model not containing sire RBV.

At a constant IMF RBV, progeny from PD rams had 0.22±0.092 units more IMF than lambs from WS sires (P=0.015; Table 1). There was no interaction between IMF RBV and breed, indicating that the effect of IMF RBV on IMF of the progeny is the same across the two terminal breeds.

Sex, PDS and HCWT had a significant effect on IMF (P<0.001; Table 1). Female lambs had 0.17 ± 0.041% units more IMF than males. As the HCWT of the lamb increased from 18kg to 30kg, IMF increased from 3.8 ± 0.08% to 4.5 ± 0.10%. Therefore, for every 1kg increase in HCWT, there was a 0.06 unit increase in IMF. These effects are similar to those reported for the Information Nucleus Flock (Pannier *et al.*, 2014).

Table 1. Degrees of freedom (number [NDF]; and denomimator [DDF]), F value and probabilites of the fixed effects in the mixed model for IMF and Shear Force

		IMF			Shear force	
	NDF, DDF	**F Value**	**Pr > F**	**NDF, DDF**	**F Value**	**Pr > F**
RBV	1, 1167	35.53	<0.001	1, 1125	21.26	<0.001
Breed	1, 1167	5.99	0.015	1, 1125	0.27	n.s.
FARM	15, 1167	15.13	<0.001	15, 1125	3.16	<0.001
SEX	1, 1167	17.59	<0.001	1, 1125	5.2	0.023
SLDATE(FARM)	6, 1167	1.37	n.s	6, 1125	15.1	<0.001
BT(FARM)	12, 1167	1.68	n.s.	12, 1125	1.44	n.s.
FARM*SEX				15, 1125	1.74	0.038
HCWT	1, 1167	35.17	<0.001	1, 1125	7.95	0.005
HCWT*HCWT				1, 1125	6.57	0.011
HCWT*FARM				15, 1125	2.65	<0.001
HCWT*SEX				1, 1125	4.28	0.039

Relationship between SF5 RBV and progeny performance. Sire was a marginally significant covariate for shear force (P=0.052). When SF5 RBV was included as a covariate, the RBV had a significant effect on progeny shear force (P<0.001). Across a 10N SF5 RBV range,

progeny shear force increased by 7.2 N of shear force, resulting in a 0.7 ± 0.16N increase in shear force associated with 1N increase in SF5 RBV (Figure 1b). There was no effect of breed, nor any interaction between breed and SF5 RBV (Table 1).

PDS, sex and slaughter date within PDS had a significant impact on shear force ($P<0.001$; Table 1). Female lambs (35.9 ± 0.71N) were more tender than male lambs (37.4 ± 0.65N; $P=0.023$). HCWT and HCWT*HCWT were significant covariates for shear force ($P<0.001$; Table 1). As the HCWT of the lamb increased from 18kg to 30kg, shear force changed from 41.8 ± 1.38N to 37.1 ± 2.4N.

SF5 RBV had a significant effect on LMY ($P<0.001$); a 1 N decrease in SF5 RBV resulted in 0.1 ± 0.03% decrease in progeny LMY. There was a 1.9% range in LMY sire solutions across the dataset examined, and 11.7N range in sire SF5 RBVs. This means that producers of terminal sired lambs selecting rams based solely on SF5 are likely to decrease LMY in their lambs.

IMF RBV had a significant effect on SF5 ($P=0.003$); a 1% increase in IMF RBV resulted in a 3.3 ± 1.10N decline in shear force. Similarly, a 1N decrease in SF5 RBV resulted in a 0.08 ± 0.016% increase in IMF ($P<0.0001$). This means selecting for either SF5 or IMF will have a positive effect on both eating quality traits in terminal lambs.

CONCLUSION

The newly created RBVs for EQ traits for terminal sires have a significant linear impact on the phenotypic expression in their progeny for IMF and SF5. Use of terminal rams with desirable IMF and SF5 RBVs will provide lamb consumers with better eating quality product. Both eating quality and other carcase traits need to be taken into consideration simultaneously in genetic improvement programs in terminal lamb productions systems.

ACKNOWLEDGEMENTS

The LMY & EQ PDS project was supported by Australian lamb producers through Meat & Livestock Australia. The contribution of 20 site hosts, 60 ram breeders, 9 site facilitators, 3 lamb supply chain officers, 7 supply chains, 13 processing plants, 4 slaughter measurement teams, 5 laboratories and Sheep Genetics is greatly appreciated.

REFERENCES

Daetwyler, H., Swan, A., van der Werf, J. & Hayes, B. (2012). *Genet. Select. Evol.* **44**: 33.

Hopkins, D. L., Hegarty, R. S. & Farrell, T. C. (2005). *Aust. J. Exp. Agric.* **45**: 525.

Hopkins, D. L., Stanley, D. F., Toohey, E. S., Gardner, G. E., Pethick, D. W. & van de Ven, R. (2007). *Aust. J. Exp. Agric.* **47**: 1219.

Mortimer, S. I., van der Werf, J. H. J., Jacob, R. H., Hopkins, D. L., Pannier, L., Pearce, K. L., Gardner, G. E., Warner, R. D., Geesink, G. H., Hocking Edwards, J. E., Ponnampalam, E. N., Ball, A. J., Gilmour, A. R. & Pethick, D. W. (2014). *Meat Science* **96**(B): 1016.

Pannier, L., Pethick, D. W., Geesink, G. H., Ball, A. J., Jacob, R. H. & Gardner, G. E. (2014). *Meat Science* **96**(B): 1068.

Pearce, K. L. (2009). Sheep CRC program 3: Next generation meat quality project 3.1.1 phenotyping the information nucleus flocks: Operational protocol series., Vol. 1, pp 107 Perth, Western Australia: Murdoch University.

Pethick, D. W., Ball, A. J., Banks, R. G. & Hocquette, J. F. (2011). *Anim. Prod. Sci.* **51**: 13.

Proc. Assoc. Advmt. Breed. Genet. 21: 145-148

GENOMIC BEST LINEAR UNBIASED PREDICTION USING DIFFERENTIAL EVOLUTION

H.A. Al-Mamun[1], P. Kwan[2], S. Clark[1], S.H. Lee[3], H.K. Lee[3], K.D. Song[3], S.H. Lee[4] and C. Gondro[1]

School of Environmental and Rural Science, University of New England, NSW, Australia
[2]School of Science and Technology, University of New England, NSW, Australia
[3]The Animal Genomics and Breeding Center, Hankyong National University, Anseong, Korea
[4]Division of Animal and Dairy science, Chung Nam National University, Daejeon, Korea

SUMMARY

In this paper we proposed a method to improve the accuracy of prediction of genomic best linear unbiased prediction (GBLUP). In GBLUP a genomic relationship matrix (GRM) is used to define the variance-covariance relationship between individuals and is calculated from all available genotyped markers. Instead of using all markers to build the GRM, which is then used for trait prediction, we used an evolutionary algorithm (differential evolution – DE) to subset the marker set and identify the markers that best capture the variance-covariance structure between individuals for specific traits. This subset of markers was then used to build a trait relationship matrix (TRM) that replaces the GRM in GBLUP (herein referred to as TBLUP). The predictive ability of TBLUP was compared against GBLUP and a Bayesian method (Bayesian LASSO) using simulated and real data. We found that TBLUP has better predictive ability than GBLUP and Bayesian LASSO in almost all scenarios.

INTRODUCTION

Genomic selection is a method based on marker-assisted selection that is used to determine the genetic value of individuals so that they can be selected as parents in breeding programs. In genomic selection, marker effects are estimated from a *discovery* (or training) dataset that comprises individuals that have both genotypic and phenotypic information. Then genomic estimated breeding values (GEBV) for selection candidates without phenotypic records are estimated based on these marker effects. Within the framework of genomic selection, two different approaches are commonly used to estimate the marker effects in the training data. The first approach assumes all SNP have a non-zero contribution to the variance of the trait of interest and the distribution of the SNP effects follows a normal distribution. Both ridge regression best linear unbiased prediction (RR-BLUP) and genomic best linear unbiased prediction (GBLUP) are based on this assumption. The second approach is based on non-linear methods that emphasize certain genomic regions and allow marker effects to come from different statistical distributions. Bayes A, Bayes B (Meuwissen *et al.* 2001), Bayes C (Habier *et al.* 2011) and Bayesian LASSO (Least Absolute Shrinkage and Selection Operator) (de los Campos *et al.* 2009) are examples of such non-linear methods for genomic selection.

GBLUP was first suggested by VanRaden (VanRaden 2008) and has been used for prediction of breeding values for use in agricultural selection programs (Goddard & Hayes 2009). In GBLUP a genomic relationship matrix (GRM) is used to define the variance-covariance relationship between individuals and is calculated from all available genotyped markers. Most of the proposed (VanRaden 2008; Goddard *et al.* 2011) implementations of the GRM are based on the infinitesimal model which assumes that a very large number of genes are evenly distributed across the genome, each contributing a minute amount to the trait of interest. In GBLUP the same GRM is used for the estimation of GEBV irrespective of the trait. Most traits of interest in animal or plant breeding are in fact polygenic but not necessarily infinitesimal; i.e. different traits are

controlled by (a limited) different sets of genes. The true underlying genetic structure of any trait deviates from the infinitesimal model to a certain extent and most quantitative traits are significantly affected by a finite set of genes (Meuwissen et al. 2001). Therefore, a GRM estimated based on the assumption of the infinitesimal model cannot optimally describe the variance-covariance relationships between individuals for the trait of interest. A model that uses only the SNP that track the relevant regions (QTL) of the traits of interest may be more appropriate to construct the variance-covariance relationship matrix.

Whereas methods that place different weightings on markers have also been proposed (i.e. Bayes A, B, C, and R), studies in which evolutionary algorithms like Differential Evolution (Storn & Price 1995) were applied to solve such a problem are few. Differential Evolution (DE) is a reliable and versatile function optimizer that is easy to implement, fast to converge, and does not require complex initial settings. DE has been successfully used in a wide range of biological optimization problems. The objective of this study was to apply DE in identifying an optimum subset of SNP to construct the variance-covariance relationship matrix for a specific trait, followed by estimation of the GEBV using BLUP based on this matrix. The performance of this method, called *Trait Best Linear Unbiased Prediction* (TBLUP) was assessed by comparing it with GBLUP and a Bayesian method (Bayesian LASSO) on simulated and real data.

MATERIALS AND METHODS

Data. One real dataset and one simulated dataset were used to assess the proposed method. Genotype information on 50K Illumina BeadChip array was available for a total of 1,937 cattle from pure-breed Korean Hanwoo with four phenotypic data: back fat (BF), carcass weight (CW) eye muscle area (EMA), and marbling score (MS). The simulated dataset contains genotype information on 10,000 samples for 40,000 SNP with simulated phenotypes. Genotypes were simulated by random sampling from frequencies under Hardy-Weinberg equilibrium (in effect an unstructured population). Phenotypes were simulated for different numbers (50, 100, 200 and 500) of known QTL. Randomly selected SNP were assigned different effects drawn from a normal distribution. Then the phenotypes were created by summing up the SNP effects plus a random environmental effect component. Both the real and the simulated datasets were divided into discovery and validation populations: 100 samples were randomly selected as validation samples; the remainder of the data were used as the discovery population. The 100 random samples selected for validation were the same for all scenarios.

Evolutionary algorithm. An algorithm based on DE was developed to select the best SNP subset in order to create the genomic relationship matrix (GRM). To select a SNP subset for the GRM, random keys were used. A random key is an evolvable vector of real values (one for each SNP) that are sorted by the objective function. The ranking of the key is then used to rank the SNP. The idea is that SNP that are better for genomic prediction evolve to higher values in the key with the rest to lower values. Once the keys are sorted, they reflect the relative value of a given SNP. An additional parameter to be optimized is the number of SNP in the panel – a *cutoff value*. The DE evolves the cutoff value, sorts the SNP based on their key values and uses the top ranked ones up to the number defined by the cutoff value. More in-depth details on the algorithm are given in (Gondro & Kwan 2012). An objective function was used to calculate the fitness of the selected SNP. In the objective function, the discovery population was further divided into two subsets: i) a subset population with known phenotype, and ii) another subset population with unknown phenotype (phenotypes were set to missing for these samples). A genomic relationship matrix was constructed using only the selected SNP for all discovery samples, which was then used to predict (by using GBLUP) the phenotype for the samples in the unknown subset population. The fitness of a selected SNP subset was defined as the correlation between the actual phenotype and the predicted phenotype. For each phenotype the DE evolved for 1,000 generations.

RESULTS AND DISCUSSION

Figure 1 shows a comparison of true genetic value (TGV) vs predicted breeding value for 50 known QTLs. Table 1 shows the comparison between prediction accuracies for 50, 100, 200 and 500 known QTLs with the simulated dataset for the different methods of genomic prediction (GBLUP, Bayesian LASSO, and our proposed method TBLUP). For the simulated data, the proposed method performed better than GBLUP and Bayesian LASSO. For the real dataset, heritability of the phenotypes were estimated using the GCTA software (Yang *et al.* 2011) which were 0.54, 0.56, 0.53 and 0.43 for BF, CW, EMA and MS respectively. Table 2 shows the genomic prediction accuracies for the validation samples obtained in the real data achieved by the three methods. Once again, the proposed method outperformed the two other methods for all four phenotypes.

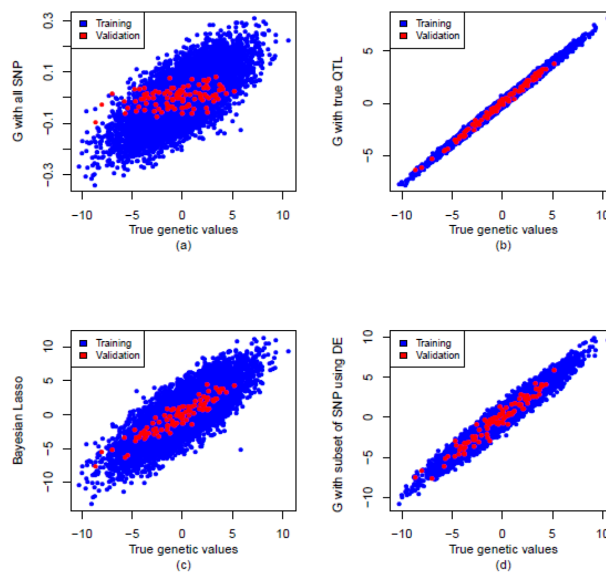


Figure 1. Prediction accuracy comparison (simulated phenotype with 50 known QTL). Blue dots are predicted values for the training data while the red dots are predicted values for the validation data. (a) Accuracy using all SNP, (b) accuracy using only the *true* QTL (QTN), (c) accuracy using Bayesian Lasso (BLR), and (d) accuracy using DE.

Table 1. Prediction accuracy comparison with the simulated data

True QTL	GBLUP	BL*	TBLUP	
			Accuracy	SNP used / QTL found
50	0.40	0.94	0.97	111 / 36
100	0.32	0.89	0.96	172 / 62
200	0.33	0.83	0.98	469 / 107
500	0.20	0.69	0.95	1041 / 186

Table 2. Prediction accuracy comparison with the real data

Trait	GBLUP	TBLUP	BL*
BF	0.370	0.440	0.394
CW	0.350	0.416	0.263
EMA	0.355	0.410	0.325
MS	0.236	0.245	0.233

*Bayesian LASSO

Improved accuracy of genomic prediction has immediate practical and commercial value for agricultural production as it leads to improved accuracy of selection and higher rates of genetic gain. GBLUP and various Bayesian methods for genomic prediction have been successfully employed in a large number of scenarios. The accuracy of these genomic predictions depends on

the genetic architecture of the trait, e.g. number of QTL and their effect sizes (Hayes *et al.* 2010), marker density, linkage disequilibrium (LD) and family relationships (Goddard *et al.* 2011; Clark *et al.* 2012; Wientjes *et al.* 2013), population structure (Moghaddar *et al.* 2014), sample size (Goddard 2009) and also the method used to estimate marker effects (Clark *et al.* 2011). Bayesian methods tend to outperform BLUP approaches when the trait is less polygenic (Clark *et al.* 2011). In practice, differences between methods in prediction accuracy are generally quite small. While these methods have well characterised statistical properties they are constrained by the underlying model assumptions. Given the dimensionality of the solution space, even very small estimates of effects in non-informative markers (noise) will, collectively, reduce prediction accuracy. This is an increasing problem with the increasing number of genetic variants to predict from. In TBLUP, we have attempted to reduce the noise from the system and tried to identify only the SNP that tracked relevant regions. In essence, the approach attempts to create a relationship matrix that tracks relationships between causal regions while excluding spurious associations and even true genetic relatedness that is not relevant to the trait of interest. We suggest that a *model free heuristic optimisation* approach choosing a small subset of best predictors is expected (and shown in the present study) to perform better in the context of genomic prediction.

CONCLUSION

In summary, we have described a novel BLUP method for estimation of breeding values using a trait-based relationship matrix, which we called TBLUP. The only difference between conventional GBLUP and the proposed TBLUP is that TBLUP focuses more closely on those markers that effectively contribute to the variation of the trait of interest and removes some of the noise that reduces accuracy of prediction. The preliminary results with real data were promising but further sudies (with more real data) are required to properly validate the method and better understand its advantages and limitations. In practice, the method can be used to develop smaller panel sets and this should reduce genotyping costs which can lead to a wider adoption by industry.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ01134906), Rural Development Administration, Republic of Korea and an Australian Research Council Discovery Project DP130100542.

REFERENCES

Clark S.A., Hickey J.M. & van der Werf J.H.J. (2011) *Genet. Sel. Evol.* **43**: 18.
Clark S.A., Hickey J.M., Daetwyler H.D. & van der Werf J.H.J. (2012). *Genet. Sel. Evol.* **44**: 4.
de los Campos G., Naya H., Gianola D., Crossa J., Legarra A., Manfredi E., Weigel K. & Cotes J.M. (2009) *Genetics* **182**: 375-85.
Goddard M. (2009) Genetica **136**: 245-57.
Goddard M.E. & Hayes B.J. (2009) *Nat. Rev. Genet.* **10**: 381-91.
Goddard M.E., Hayes B.J. & Meuwissen T.H.E. (2011) *J. Anim. Breed. Genet.* **128**: 409-21.
Gondro C. & Kwan P. (2012). IGI Global.
Habier D., Fernando R.L., Kizilkaya K. & Garrick D.J. (2011) *BMC Bioinformatics* **12**: 186.
Hayes B.J., Pryce J., Chamberlain A.J., Bowman P.J. & Goddard M.E. (2010) *PLoS Genet*. **6**.
Meuwissen T.H., Hayes B.J. & Goddard M.E. (2001) *Genetics* **157**: 1819-29.
Moghaddar N., Swan A.A. & van der Werf J.H.J. (2014) Genet. Sel. Evol. **46**: 58.
Storn R. & Price K. (1995) In: *Technical Report, TR-95-012* Int. Comp. Sci. Inst. (ICSI).
VanRaden P.M. (2008) *J. dairy sci.* **91**: 4414-23.
Wientjes Y.C.J., Veerkamp R.F. & Calus M.P.L. (2013) Genetics **193**: 621-31.
Yang J., Lee S.H., Goddard M.E. & Visscher P.M. (2011) *Am. J. Hum. Genet.* **88**: 76-82.

Proc. Assoc. Advmt. Breed. Genet. 21: 149-152

IMPROVING THE ACCURACY OF ACROSS BREED GENOMIC PREDICTIONS

M.E. Goddard[1,3], I.M. MacLeod[1,2,3], S. Bolormaa[3], B.J. Hayes[3,4], and K.E. Kemper[1]

[1] Faculty of Veterinary & Agricultural Science, University of Melbourne, Victoria, Australia
[2] Dairy Futures Cooperative Research Centre, La Trobe University, Bundoora VIC 3083, Australia
[3] AgriBio, Dept. Economic Dev., Jobs, Transport & Resources, Victoria, Australia
[4] Biosciences Research Centre, La Trobe University, Victoria, Australia

SUMMARY

Genomic predictions derived from one breed but applied in another breed typically have low accuracy due to SNP x breed interactions (due to either QTL x breed interactions or differences in LD between breeds) and due to differences between breeds in QTL allele frequency. In this paper we discuss the importance of these two factors and the implications for livestock breeding.

INTRODUCTION

Genomic selection (Meuwissen *et al.* 2001) has been very successful in predicting the breeding value of animals from DNA marker data. It works best in Holstein cattle for milk production traits where there is a large amount of high quality data on which to train the prediction and the animals where the prediction is to be used (target animals) are closely related to the training population (often their sons). As the test animals become less closely related to the training population the accuracy of prediction declines (Habier et al 2010) and when they are of different breeds, the accuracy is typically low (eg. Kemper et al 2015a). Unfortunately there are many traits where we do not have a large training dataset within every breed. For instance, the expense of measuring feed conversion efficiency limits the size of training datasets. Therefore we would like to use a multi-breed training dataset to maximize the number of training animals and to predict breeding value in animals of a breed included in the training data or even a breed not included in the training data. In these situations the low accuracy of across breed prediction is a severe disadvantage. Alternately, if a method was available in which across breed prediction was of high accuracy, it seems likely that within breed prediction would also be more robust and not affected by the degree of relationship between training and test animals. In this paper we consider reasons for the low accuracy of across breed prediction and what might be done to increase the accuracy.

ACROSS BREED ACCURACY OF GENOMIC PREDICTIONS

Information from another breed can be used in prediction in two ways. Firstly, Brondum *et al.* (2012) and Khansefid *et al.* (2014) showed that the accuracy of prediction in breed B could be increased by including in the statistical model SNPs that were associated with the trait in breed A but by estimating the effect of the SNP entirely within breed B. This implies that some of the same QTL segregate in both breeds. Secondly, the accuracy can be improved slightly by estimating the effect of each SNP across all target breeds (Bolormaa *et al.* 2013a, Hoze *et al.* 2014, Makgahlele *et al.* 2013) and some accuracy is obtained even in a breed not included in the training population (Kemper *et al.* 2015a). For instance, Kemper *et al.* (2015a) found the accuracy in Australian Red cattle for milk production traits averaged 0.3 based on a training population of Holstein and Jerseys. They also reported that the regression of phenotype on genomic EBV (bias) was 0.6 on average indicating that the EBVs exaggerated the predicted differences between animals.

REASONS FOR LOW ACCURACY

There are two reasons for low across breed prediction accuracies – SNP x breed interactions and differences in QTL allele frequency between breeds. Khansefid *et al.* (2014) analysed residual

feed intake using genomic relationship matrices and found the SNP variance and the SNP x breed variance to be about equal. SNP x breed interactions could be due to QTL x breed interactions or differences between breeds in linkage disequilibrium (LD) between QTL and SNPs.

The extent of QTL x breed interactions, due to non-additive gene effects, is largely unknown. It is similar to sire by breed interaction which is seldom above 0.2 of the genetic variance, so it seems unlikely that QTL x breed would explain 0.5 of the genetic variance. One type of non-additive variance is dominance which Bolormaa *et al.* (2015) estimates to explain 5% of phenotypic variance across a number of traits in beef cattle.

Differences in LD between breeds depend on the distance which separates the QTL and SNP. GBLUP uses LD over long distances within a breed but this LD breaks down between breeds (deRoos *et al.* 2008). Therefore this will explain some of the SNP x breed interaction found by Khansefid *et al.* (2014). Differences in LD occur even at short distances in the case of a QTL mutation which has occurred since the breeds diverged. In this case, an ancestral haplotype may exist in one breed with the ancestral QTL allele and in the other breed with the mutant QTL allele. Thus even with sequence data and methods such as BayesR (Erbe *et al*. 2012), SNP x breed interaction will occur for recent QTL mutations unless the QTL mutation itself is used.

Differences in QTL allele frequency will occur due to selection and drift. QTL with minor allele frequency (MAF) near to 0.5 contribute more to genetic variance and have their effect estimated more accurately than QTL with low MAF. This increases the accuracy of prediction within a breed. However, if the MAF is low in the training population but high in the target population, the QTL is important to genetic variance in the target population but its effect will be estimated poorly. This will reduce the accuracy of the genomic prediction in the target population. The extreme case of this phenomenon is when the QTL segregates in the target population but not in the training population. In that case, no estimate of its effect can be made and the variance it explains will be totally missed. This can happen if the QTL mutation is recent and has occurred in one breed since the breeds diverged. It can also occur if the QTL is old but has become fixed in the training population but not the target population. This extreme case, where the QTL segregates only in one breed, places an upper limit on the potential accuracy attainable using across breed genomic prediction. Kemper *et al.* (2015b) recently estimated that about ½ the QTL discovered for milk production traits in Holstein also segregated in Jersey cattle. Even among random SNPs in sequence data, 20% do not segregate in both Holsteins and Jerseys (Kemper et al 2015b).

The factors that cause low accuracy of across breed predictions can also cause bias. For instance, if a QTL mutation has occurred in the training population since the breeds diverged, the genomic predictions will predict that the QTL contributes to variance in the target population when it does not segregate. Kemper *et al.* (2015b) documented several examples where Holstein-only QTL were predicted for Jersey cattle when the QTL did not segregate in that breed.

AGE AND BREED DISTRIBUTION OF QTL

From the discussion of factors causing low accuracy it emerges that two closely related parameters are important - the age of QTL mutations and the range of breeds in which they segregate. Since the mutation causing a QTL (a QTN) is seldom known, the age and distribution of QTL is not well understood. For neutral mutations we can estimate their average age from the ratio of the heterozygosity per site (in cattle this is typically about 0.001) to the heterozygosity introduced each generation by mutation (2 x mutation rate e.g. 2 x 10^{-8}) which gives an average age of 50,000 generations or well before domestication of cattle and sheep. This average disguises a large range in age from new mutations in the last generation to very old ones.

QTL are unlikely to be neutral and so selection will modify this average age. One estimate of average age is the genetic variance (e.g. $0.5V_E$) divided by the variance added by mutation each generation (e.g. $0.001V_E$) or about 500 generations. However, this average includes detrimental

mutations of large effect which are soon eliminated from the population, so the average age of those that are segregating is likely to be much greater than 500 generations and to vary greatly about this average. The age of QTL can also be estimated from the length of a common haplotype which surrounds the mutant allele. For myostatin mutations causing double muscling O'Rourke *et al*. (2010) estimated their age to be <100 generations. Consistent with this, each mutation segregates in one or a few related breeds.

Kemper *et al*. (2015b) examined the length of haplotype surrounding QTL for milk traits in Holstein. They found examples of QTL that appeared recent (800 generations) and occurred in Holsteins but not Jersey and others that appeared old (12000 generations) and occurred in both breeds. By comparison de Roos *et al*. (2008) estimated the age of the Holstein-Jersey divergence at 400 generations. There were also QTL that appeared old but did not segregate in Jerseys. This is expected to occur because breeds of *Bos taurus* cattle have suffered some inbreeding since they diverged and consequently lost a proportion of polymorphisms including QTL. Of 11 QTL in Holstein, 6 also segregated in Jerseys.

Saatchi *et al.* (2014) found 4 QTL for weight that segregated in several breeds of US beef cattle and we have found QTL in the same position in Australian beef cattle. On the other hand, Bolormaa *et al.* (2013b) concluded the QTL seldom segregate in both *B. taurus* and *B. indicus,* which diverged perhaps 100,000 generations ago.

Thus we conclude that while many QTL segregate in multiple *B. taurus* breeds, some QTL segregate only in some breeds, either because they are recent mutations or because they are old but fixed in some breeds.

IMPLICATIONS FOR LIVESTOCK BREEDING

We conclude that the poor accuracy of genomic prediction when the training dataset comes from one breed and the predictions are applied to another breed is due to a combination of QTL x breed interactions, differences in LD between breeds and differences in QTL allele frequency between breeds (especially when a QTL segregates only in some breeds). What strategies can be used to overcome this problem?

One strategy is to do all prediction within breed. This is simple to implement because a low density SNP panel (e.g. 50K) is satisfactory and simple statistical methods, such as BLUP, can be used. However, this strategy cannot be implemented in all cases and even where it can be used it has disadvantages – some 10% of the genetic variance is not explained by a 50K SNP panel and the predictions may not be robust when applied to target animals not closely related to the training population (MacLeod *et al.* 2014a).

The alternative strategy is to use a multi-breed training population. This requires dense SNPs, ideally sequence data, and a statistical method which can find and utilise the causal mutations or markers in near complete LD with them. Figure 1 shows results for a mixed breed reference (Holstein and Jersey) where accuracy was evaluated in 3 validation sets that differed in their relatedness to the reference. We compared accuracy for different SNP densities: including a combined set of high density markers and imputed sequence in and near coding regions (SEQ). Instead of real phenotypes, we simulated 4000 QTL into the real genotypes (h^2=0.6) so that the true breeding value was known. We found accuracy increased as the density of SNP increased to SEQ and this was most apparent in the target population that was least related to the training population (Australian Red). However, there was still a drop in accuracy even with the QTL included in the SEQ data. This must be partly a result of the BayesR analysis spreading the effect of a single QTL across several SNP in strong LD with the QTL in the training population. Figure 1 also demonstrates that the BayesRC method (Macleod *et al.* 2014b), which is similar to BayesR but includes a broadly defined biological prior, also increases the accuracy of across breed genomic predictions.

As causal mutations or markers in near perfect LD with causal mutations are discovered we will be better able to assess the importance of non-additive genetic effects causing QTL x breed interactions and fit them in the model if necessary.

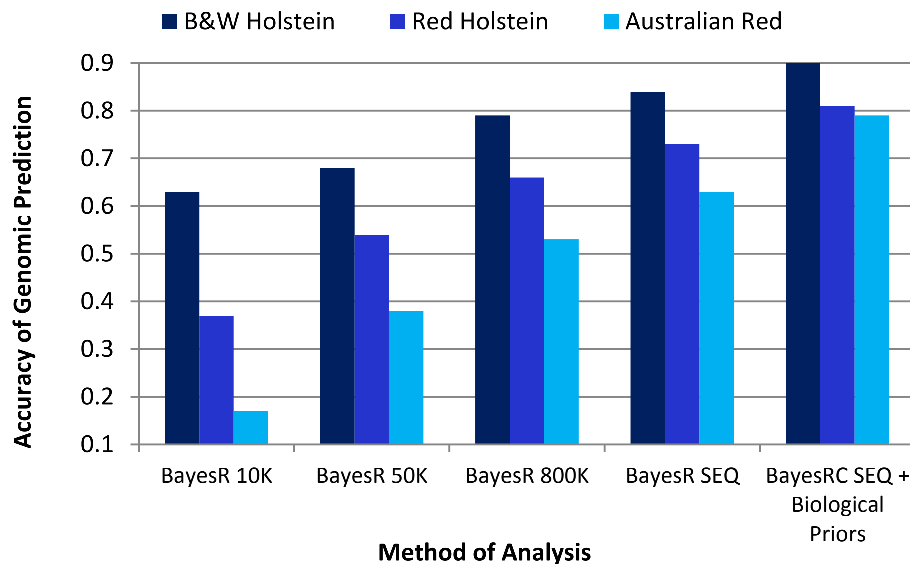


Figure 1. Accuracy of genomic prediction for simulated QTL using different densities of genotypes [(10K, 50K, 800K SNP or including sequence variants (SEQ)] with BayesR or BayesRC. Validation populations were either closely related to the reference (Black & White [B&W] Holstein), somewhat related (Red Holstein) or a different breed (Australian Red)

REFERENCES

Brondum, R.F., Su, G., Lund, M.S., Bowman, P.J. *et al.* (2012) *BMC Genomics* **13**:543.
Bolormaa, S., Pryce, J.E., Kemper, K.E., Savin, K. *et al.* (2013a) *J Anim Sci* **91:**3088.
Bolormaa, S., Pryce, J.E., Kemper, K.E., Hayes, B.J. *et al.*(2013b) *Gen Sel Evol* **45:**43.
Bolormaa, S. Pryce, J.E., Zhang, Y., Reverter, A. *et al.* (2015) *Gen Sel Evol* **47:**26.
de Roos, A.P.W., Hayes, B.J., Spelman, R.J. and Goddard, M.E. (2008) *Genetics* **179:**1503.
Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S. *et al.* (2012) *J Dairy Sci* **95:**4114.
Habier, D., Tetens, J., Seefried, F. R., Lichtner, P., & Thaller, G. (2010). *Gen Sel Evol*. **42**:5.
Hozé C, Fritz S, Phocas F, Boichard D., *et al.* (2014) *J Dairy Sci* **97**:3918.
Kemper, K.E. Reich, C.M., Bowman, P.J., vander Jagt, C.J. *et al.* (2015a) *Gen Sel Evol* **47:**29.
Kemper, K.E. Hayes, B.J., Daetwyler, H.D. and Goddard, M.E. (2015b) *J Anim Breed Genet* **132:**121.
Khansefid, M., Pryce, J.E., Bolormaa, S., Miller, S.P. *et al.* (2014) *J Anim Sci* **92:**3270.
MacLeod, I.M., Hayes, B.J., Goddard, M.E. (2014a). *Genetics* **198:**1671.
MacLeod, I.M., Hayes, B.J., Vander Jagt, C.J., Kemper, K.E. *et al.* (2014b). *Proc. 10th World Congress Genetics Applied to Livestock Production*
Makgahlela ML, Mäntysaari EA, Strandén I, *et al.* (2013) *J Anim Breed Genet* **130:**10.
Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001) *Genetics* **157:**1819.
O'Rourke, B.A., Greenwood, P.L., Arthur, P.F. and Goddard, M.E. (2013). *Anim. Genet* **44:**86.
Saatchi, M., Schnabel, R., Taylor, J. and Garrick, D. (2014) *BMC Genomics* **15:**442.

Proc. Assoc. Advmt. Breed. Genet. 21: 153-156

DATA COMPRESSION: A NEW WAY TO INFER GENOMIC RELATIONSHIP MATRICES AND HIGHLIGHT REGIONS OF INTEREST IN COMMERCIAL LINES OF BROILER CHICKEN

N.J. Hudson[1], R. Hawken[2] R. Sapp[2] and A. Reverter[1]

[1] CSIRO Agriculture Flagship, Brisbane, Australia
[2] Cobb Vantress, Arkansas, USA

SUMMARY

Gene discovery relies on knowledge of animal relatedness. This in turn exploits correlation based measures of similarity now based on shared patterns of genome-wide single nucleotide polymorphism (SNP) genotypes. These comparisons are captured by the genomic relationship matrix (GRM). However, it is not clear whether correlation is the best way of quantifying those shared patterns. Here, we continue our exploration of whether one can build relationship matrices based on the concept of compression efficiency from Information Theory. Drawing on 4 commercial broiler lines, 2 lines based on growth and efficiency selected roosters, and 2 lines based on reproductive performance selected hens, we found that data compression clustered the lines by gender. Further, a sliding window version of the approach identified different gene regions apparently selected in male versus female lines. In males two prominent regions harboured *IGF-1* (Chromosome 1) and a cognate *IGF-1* receptor *INSR* (Chromosome 28). In the female lines, the reproductive hormone receptor *GNRHR* (Chromosome 10) and folate metabolism *FOLH1* (Chromosome 1) were prioritised.

INTRODUCTION

Gene discovery through genome-wide association studies (GWAS) and identification of signatures of selection require that population structure and relatedness can first be quantified and subsequently accounted for. Genetic relatedness is currently estimated by a combination of traditional pedigree-based approaches (Henderson 1975) and, given the recent availability of molecular information, the use of marker genotypes via the genomic relationship matrix (GRM) (Van Raden 2008). To date, GRM from SNP genotypes are essentially estimated using correlation.

Here, we continue our exploration as to whether the concept of compression efficiency from Information Theory can provide a complementary method for establishing patterns of genetic relatedness. The basic principle of Normalised Compression Distance (NCD) (Cilibrasi and Vitanyi 2005) is that if patterns of data in one genotype file can be used to compress shared patterns of data in the second genotype file, the two genotypes are considered related. Consequently, a short distance (high similarity) will be awarded. This process can be repeated across a genotyped population of animals to build a Compression Relationship Matrix (CRM) analogous to a GRM. This concept has previously been used by our group in both sheep and cattle populations where we have found that the NCD method can sensitively discriminate sire groups, breeds and indeed half-sibs from full sibs, in circumstances where GRM could not (Hudson *et al.* 2014 WCGALP). Moreover, we found CRM explained more genetic variance, reduced the missing heritability and yielded higher phenotype accuracies than GRM (*unpublished data*). Additionally, a preliminary version of the approach was able to cluster individual humans by ethnic group in a manner consistent with F_{ST} and known phylogeography (Hudson *et al*. 2014).

In this exploratory paper we assess the application of NCD to patterns of relatedness between 4 commercial lines of broiler chickens, *Gallus gallus domesticus*. We also use a genome-wide sliding window based on compression efficiency to identify possible signatures of selection present on a gender-specific basis.

MATERIALS AND METHODS

Populations and data resources. We used data from 988 chickens from 4 commercial lines of broilers – hereon in denoted as Lines A, B, C and D (Table 1). Individuals were selected from a much larger population of over 50,000 birds and based on full sib families to a near-balanced design of ~250 individuals per line.

Table 1. Summary of the 4 chicken lines used for this analysis

Line	Selection	Birds	Full-Sib Families	Females	Males
A	Female	204	14	167	37
B	Female	244	5	153	91
C	Male	254	18	195	59
D	Male	286	50	220	66

Two of the lines (A and B) are lines that have been generated for selecting genetically superior females – the selection focus being primarily on desirable reproductive traits. For male lines (C and D), the selection foci have been growth rate, muscle mass and feed efficiency. All animals were genotyped for 51,713 SNP (Groenen *et al.* 2009) distributed genome-wide.

Population clustering. We used NCD to compare pairs of individuals (x and y) from all 4 lines based on their respective SNP genotypes as follows:

$$NCD(x, y) = \frac{Z(xy) - \min\{Z(x), Z(y)\}}{\max\{Z(x), Z(y)\}}$$

Z(xy) represents the size of the compressed file containing both concatenated SNP genotype sequences to be compared and *Z(x)* and *Z(y)* is the size of the compressed file with the isolated SNP genotypes for *x* and *y*, respectively. We used GZIP to perform the data compression.

Signatures of Selection. In order to find signatures of selection and regions of evolutionary interest, we next applied a sliding window version of compression efficiency (CE) as previously described in Hudson *et al.* (2014). This approach exploits the sensitive pattern recognition capability of CE to find haplotype blocks that occur in one population but not another. In brief, the population level CE of non-overlapping windows was computed separately for the 4 broiler lines, corrected for heterozygosity (CEh). We used non-overlapping sliding windows of 100 consecutive SNP. The experimental design made use of two 'independent' lines of male and female populations, whose output could be overlaid. This approach helps improve the signal to noise ratio for identifying *bona fide* signatures of selection, against background noise emerging from population bottlenecks and other phenomena.

RESULTS AND DISCUSSION

Population clustering. Self-Self pairs (panel A) possess a GRM of close to 1, with deviations above 1 representing extent of inbreeding. GRM and NCD are both in agreement that the lines cluster by gender comparison (Panel B). Female-male line comparisons in blue are awarded a low similarity and high distance, whereas male-male and female-female line comparisons are more closely related.

Proc. Assoc. Advmt. Breed. Genet. 21: 153-156

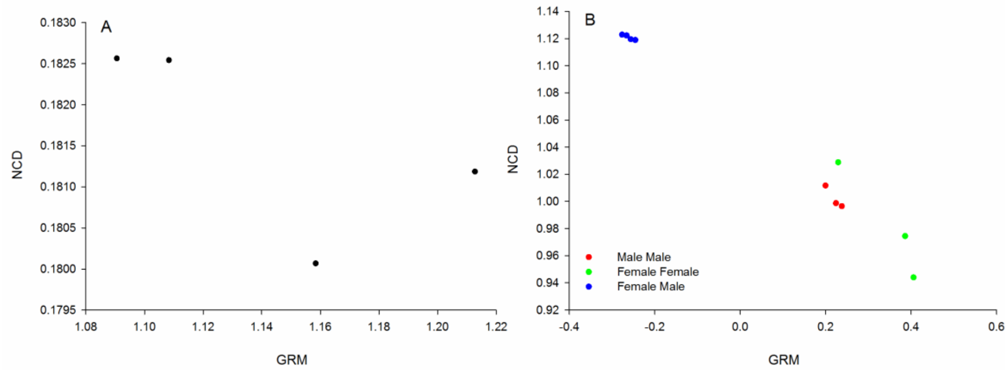


Figure 1. GRM and NCD values for the 4 chicken lines with the various gender comparisons colour coded.

Overall, there is a clear negative relationship between GRM and NCD because similarity (via correlation) is the inverse of distance (via NCD).

Signatures of selection. The genomes of all 4 lines were characterised by a large number of small peaks and a much smaller number of larger peaks. These outlier regions have particularly strong population-level scores in these regions. They would be predicted to potentially play an important role in providing the genetic basis for the phenotypes that have been selected in those populations. We manually explored the outlier regions that were gender-specific.

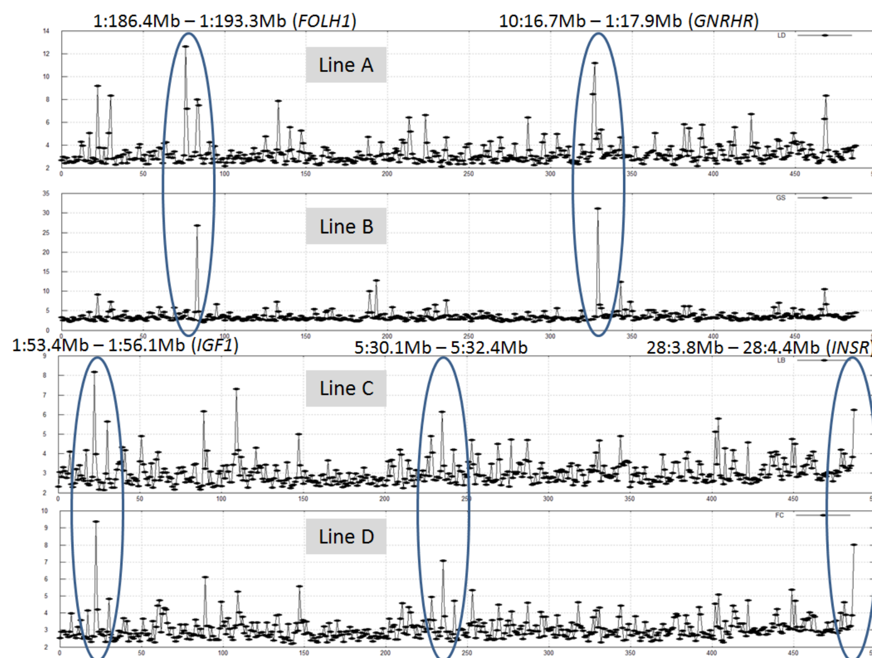


Figure 2. Compression efficiency (y-axis) of windows of 100 consecutive SNPs along the genome (x-axis) for the four chicken lines. Highlighted are the regions described in Table2.

Table 2. Regions captured by the compression efficiency of windows of 100 consecutive SNPs

Lines	Regions (Chr: Coordinates)	Example Genes in region	Total number of genes
Female	1:186.4 Mb – 1:193.3 Mb	***FOLH1***, *THRSP*	62
Female	10:16.7 Mb – 10:17.9 Mb	***GNRHR***	37
Male	1:53.4 Mb – 1:56.1 Mb	***IGF-1***, *MTERF*	45
Male	5:30.1 Mb – 5:32.4 Mb	mir-1718, mir-3532	19
Male	28:3.81 Mb – 28:4.44 Mb	***INSR***, *SIN3B*, *PEX11G*	28

In the two male lines the clear identification of two different regions containing serial components of a single functional pathway (*IGF-1* and one of its cognate receptors *INSR)* is particularly intriguing. The male lines, unlike the female lines, have been selected for increased muscle mass. IGF-1 is a well characterised master regulator of muscle mass whose molecular structure is similar to insulin. It mediates the anabolic effect of Growth Hormone (Barton 2006). This functional pairing (IGF-1 and INSR) is unlikely to occur by chance as IGF-1 is one of only three proteins to bind the insulin receptor. In an independent population of broiler chickens derived from Plymouth Rock and Cornish lines *IGF-1* had also been identified as a signature of selection (Stainton *et al.* 2015). In the female lines which have been selected for reproductive traits, we detected regions containing *GNRHR* (encoding the receptor for the reproductive hormone gonadotropin releasing hormone) and *FOLH1* (that hydrolases the vitamin folate).

Future work could fine map these genomic regions using a higher resolution (50 SNP) window, and sliding it one SNP at a time in an overlapping fashion to attempt to home in on the exact genes under selection. We have previously used this method to successfully home in on single genes across human populations, such as lactase persistence in northern Europeans and Masaai Kenyans (Hudson *et al.* 2014). The relationship matrices described in the first part of the manuscript could be 'ground-truthed' through estimation of genetic parameters, computation of missing heritability and calculation of phenotype accuracies for a phenotype of commercial interest in the broiler industry such as feed efficiency.

REFERENCES

Barton E.R. (2006). *Appl Physiol Nutr Metab* **31**:791.

Cilibrasi R. and Vitanyi P.M.B. (2005). Clustering by compression. *IEEE Trans. Inform. Theory.* **51**(12):1523.

Groenen M.A., Wahlberg P., Foglio M., Chen H.H., Megens H.J., Crooijmans R.P., Besnier F., Lathrop M., Muir W.M., Wong G.K., Gut I and Andersson L (2009). *Genome Res*. **19**:510.

Henderson C.R. (1975). Rapid method for computing the inverse of a relationship matrix. *J. Dairy Sci*. **58**(11): 1727

Hudson N.J., Porto-Neto L.R., Kijas J., McWilliam S., Taft R.J. and Reverter A. (2014). *BMC Bioinformatics* **15**:66

Hudson N.J. (2014) *WCGALP conference proceedings*. Vancouver, Canada

Stainton J.J., Haley C.S., Charlesworth B., Kranis A., Watson K. and Wiener P (2015). *Anim. Genet.* **46**: 37.

Van Raden P.M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci*. **91**:4414

Proc. Assoc. Advmt. Breed. Genet. 21: 157-160

RELATIONSHIP BETWEEN CHANGE IN TRAIT DEFINITION AND ACCURACY OF GENOMIC BREEDING VALUE OF TYPE TRAITS IN AUSTRALIAN HOLSTEIN

M. Haile-Mariam[1,2], J.E. Pryce[1,2,3] and B.J. Hayes[1,2,3]

[1]Dairy Futures Cooperative Research Centre, La Trobe University, Bundoora VIC 3083, Australia
[2]AgriBio, Dept. Economic Dev., Jobs, Transport & Resources, Bundoora VIC 3083, Australia
[3]La Trobe University, Bundoora VIC, Australia

SUMMARY

Data on type traits of Holstein cows scored between 1990 and 2014 were analysed to examine the extent of change in trait definition and to assess if these changes affect the accuracy of direct genomic values (DGV) of validation animals. The results showed that for 10 of the 20 traits, the genetic correlation of a trait that was scored before 2007 was less than 0.9 with the same trait scored after 2006. For 7 of the 10 traits, where the low genetic correlation suggested some level of change in trait definition, accuracy of their observed DGV were also markedly lower than the expected accuracy and accuracy predicted from the deterministic formula.

INTRODUCTION

Until recently type traits were not included directly in the economic index in Australia. However, from 2015, in the three new indexes (Balanced Performance Index, Health Weighted Index, Type Weighted Index), introduced by the Australian Dairy Herd Improvement Scheme (ADHIS) some type traits are included directly with their own economic values. Therefore these traits are now more important than they have been in the past. Recent results show that accuracy of direct genomic values (DGV) of type traits such as overall type and mammary system are lower than expected (Haile-Mariam *et al.* 2015). In addition, the accuracy of DGV for type traits varied considerably and explanations for such variation were not readily available (Haile-Mariam *et al.* 2013; 2015). Furthermore, results from ADHIS show that the variation in DGV among young bulls, particularly for overall type, is lower than expected. As part of the breed development program of Holstein Australia (HA) definition of some type traits has changed over time, and this may contribute to the lower than expected accuracy of DGVs for some traits.

This study explores if the low accuracies of DGVs observed for some type traits are related to changes in definitions of some type traits introduced in 2007 and we explore ways of minimising these effects on accuracy of DGVs. This is done in the following way: 1) for each type trait, estimating the genetic correlation for the trait scored on cows before 2007 and the trait scored after 2006, then assessing if the observed accuracy of DGV can be related to the genetic correlation of a trait in the two time periods; 2) for the two composite traits, namely overall type and mammary system, we assessed if predicting them from individual type traits can be used to increase genetic correlation between the two periods and thereby improve accuracy of DGVs.

MATERIALS AND METHODS

Data on Holstein cows type scored between 1990 and 2014 were obtained from the ADHIS. Cows were type scored by classifiers once during the first lactation. Details of the data used for this study is given by Haile-Mariam and Pryce (2015). The traits considered are those with relatively large amounts of data in the two periods (Table 1). The data included 18 linear traits scored on a scale of 1 to 9 and two composite traits scored on a scale of 1 to 16. In 2007 HA introduced some changes in the way type traits are scored. To assess the effect the change, data of cows scored before 2007 (period 1) and after 2006 (period 2) were split into two groups. Table 1

shows the numbers of cows scored with the mean and standard deviation for 20 traits in each period. Of the 2,724 bulls, with progeny in the second period 32% were also sires of 46% of the cows scored in the first period reflecting a good level of connectedness between the datasets.

Table 1. Number of cows classified, mean and standard deviation (SD) for type traits scored before 2007 and after 2006

Traits	Scored before 2007			Scored after 2006		
	No.	Mean	SD	No.	Mean	SD
Overall type	434,207	9.47	1.79	101,189	10.08	1.59
Mammary system	433,770	9.81	1.75	100,720	10.36	1.60
Stature	434,207	5.99	1.42	101,189	6.74	1.53
Udder texture	434,207	6.04	1.13	101,188	5.97	1.29
Bone quality	434,207	6.11	1.25	101,189	6.82	1.26
Angularity	434,207	5.46	1.14	101,189	5.84	1.10
Muzzle width	434,207	5.68	1.05	101,189	6.03	1.12
Body depth	356,080	6.25	1.12	101,189	5.59	1.32
Chest width	434,207	5.68	1.01	101,189	5.48	1.28
Pin width	434,207	5.75	1.19	101,189	6.67	1.30
Pin set	434,200	4.23	1.39	101,189	3.98	1.22
Foot angle	356,083	4.92	1.09	101,189	5.23	0.99
Rear leg set	434,203	5.47	1.01	101,189	5.28	1.07
Udder depth	356,068	5.89	1.15	101,188	5.18	1.59
Fore attachment	434,207	5.51	1.08	101,189	5.55	1.33
Rear attachment height	434,206	5.84	1.08	101,189	6.58	1.30
Rear attachment width	434,207	5.34	1.11	101,189	5.64	1.46
Central ligament	434,201	6.09	1.08	101,188	6.46	1.18
Teat placement fore	434,204	5.04	1.19	101,189	5.18	1.30
Teat length	356,025	4.44	1.46	101,188	4.55	1.35

To explore if the low DGV accuracy that we observed (e.g. Haile-Mariam *et al.* 2013; Haile-Mariam *et al.* 2015) for some traits is related to the change in the definition of type traits over time the following analyses were performed. First, to assess the extent of change in trait definition, type data of cows scored in the two time period were analysed as two different but correlated traits in a bi-variate sire model to estimate heritability (h^2) for each period and genetic correlation between the two periods. Data were analysed fitting sire as a random effect and Herd-Classifier-Round as the main fixed effect. Age and days in milk at scoring were also fitted as covariates. For overall type and mammary system, the two composite traits, value for cows scored in period 1 were predicted from other type traits that were less affected by trait definition (Table 2). A genetic correlation of below 0.90 between the two periods was considered as criteria to designate a trait whose definition changed. For both traits, a linear prediction equation based on selected linear type traits and the composite traits was developed based on data of cows scored in period 2 and applied to data of cows scored in period 1, assuming the period 2 scores as ‘gold standard’. Secondly DGV for validation bulls born in 2004 and after were predicted and accuracy was calculated as a correlation between DGV and daughter trait deviation (DTD). These DGV accuracy were adjusted using the average accuracy of the DTDs (calculated from h^2 and number of daughters) and compared to expected accuracy calculated from the prediction error variance and to the accuracy calculated using deterministic formula (e.g. Hayes *et al.* 2009). For these analyses the DTD and

genotype data of 2,407 bulls were obtained from the ADHIS. Details on the DTD, genotype and methods for genomic prediction are provided elsewhere (e.g. Haile-Mariam *et al.* 2015). The parameters used for prediction of accuracy using the deterministic formula were the same as those assumed by Hayes *et al.* (2009). However, the reference population size was 2,056 bulls for 14 of the 20 traits and 1,860 bulls for udder depth, body depth, foot angle and teat length. The h^2 in Table 2 estimated based on the data in period 1 were used to calculated effective h^2.

RESULTS AND DISCUSSION

The mean type score of cows increased from period 1 to period 2 for all traits except udder texture, body depth, chest width, pin set, rear leg set and udder depth (Table 1). The reason why the mean score of cows for udder texture and these other traits decreased could be related to the change in the way cows were scored, or to selection, if the optimum for the trait is a lower or intermediate score. For all traits, the h^2 was lower in cows scored in period 2 compared to those in period 1 (Table 2). The genetic correlation of cows scored in the two periods was very high for traits such as pin set, but was the lowest for mammary system followed by overall type (Table 2). When predicted overall type and mammary system were used for cows scored in period 1, instead of the scores by the classifiers, the genetic correlation between the two periods increased only marginally. In the case of mammary system, the correlation increased to 0.63 when 4 udder traits (udder depth, teat length, teat placement fore and fore attachment) were used as predictors. In the case of overall type, the use of 10 traits (teat length, teat placement fore, udder depth, rear set, foot angle, pin set, body depth, muzzle width, angularity, bone quality) increased the correlation to 0.68. This is slightly lower than the prediction using all type traits (0.70).

Table 2. Estimates of heritability (h^2), genetic correlation of a trait between scored before 2007 and after 2006, adjusted observed accuracy, expected accuracy and differences (Diff)

Traits	h^2 before '07	h^2 after '06	Genetic correlation	Adj. accuracy	Exp. accuracy	Diff
Overall type	0.27±0.01	0.13±0.01	0.66±0.08	0.36	0.56	0.20
Mammary system	0.27±0.01	0.14±0.01	0.59±0.07	0.33	0.55	0.22
Stature	0.45±0.01	0.30±0.02	0.84±0.03	0.46	0.61	0.15
Udder texture	0.23±0.01	0.13±0.01	0.68±0.06	0.46	0.61	0.15
Bone quality	0.30±0.01	0.25±0.01	0.97±0.02	0.54	0.64	0.10
Angularity	0.26±0.01	0.18±0.01	0.93±0.03	0.55	0.61	0.06
Muzzle width	0.23±0.01	0.20±0.01	0.90±0.03	0.53	0.63	0.10
Body depth	0.38±0.01	0.28±0.02	0.95±0.02	0.70	0.64	-0.06
Chest width	0.25±0.01	0.21±0.01	0.81±0.04	0.53	0.66	0.13
Pin width	0.35±0.01	0.26±0.02	0.85±0.04	0.57	0.64	0.07
Pin set	0.37±0.01	0.29±0.01	0.99±0.01	0.51	0.61	0.10
Foot angle	0.20±0.01	0.17±0.01	0.93±0.03	0.55	0.61	0.06
Rear leg set	0.18±0.01	0.10±0.01	0.94±0.03	0.44	0.54	0.10
Udder depth	0.40±0.01	0.33±0.02	0.93±0.02	0.71	0.65	-0.06
Fore attachment	0.22±0.01	0.16±0.01	0.88±0.04	0.42	0.61	0.19
Rear attachment height	0.27±0.01	0.16±0.01	0.71±0.06	0.52	0.59	0.07
Rear attachment width	0.23±0.01	0.18±0.01	0.69±0.06	0.59	0.62	0.03
Central ligament	0.24±0.01	0.12±0.01	0.70±0.06	0.27	0.63	0.36
Teat placement fore	0.35±0.01	0.30±0.02	0.90±0.03	0.54	0.67	0.13
Teat length	0.44±0.01	0.34±0.02	0.94±0.02	0.52	0.66	0.14

For 18 of the 20 traits analysed the observed accuracy of DGV was lower than the expected accuracy (Table 2). The observed accuracy being lower compared to the expected accuracy could be a result of change in trait definition (Table 2) over the period. In fact Table 2 shows that for 7 of the 10 traits whose genetic correlations between the two periods were < 0.9 the observed accuracy was less than the expected accuracy by at least 0.1. However, for rear attachment height and rear attachment width where the genetic correlations between period 1 and 2 were low (Table 2) the observed accuracy was only marginally lower than the expected accuracy. For teat length, pin set, rear leg set and bone quality despite having genetic correlations of > 0.9, the observed accuracy was lower than the expected accuracy by 0.1 to 0.14 (Table 2), suggesting that issues other than correlations that cover the current period may contribute to the difference.

The accuracy of prediction from the deterministic formula varied from 0.59 for foot angle (the lowest h^2) to 0.65 for stature (the highest h^2). The small variation in the formula predicted accuracy among traits is expected because all the parameters that influence accuracy, except the h^2 are the same. The highest differences between the formula predicted and the expected accuracy (0.07-0.08 in favour of the formula) were for mammary system, overall type and rear leg set. For all the other traits the difference was 0.04 or less. An additional evidence of the effect of change in trait definition detected as low genetic correlation of a trait over time on DGV accuracy was observed because the relationship of DGV accuracy with the product of h^2 and genetic correlation was stronger ($R^2 = 0.49$) than that of DGV with h^2 alone ($R^2 = 0.41$) based on all traits. Thus taking account of the genetic correlation improved agreement between observed DGV and expected accuracy.

The broader implications of these results are, when assessing the accuracy of DGV of traits that are subjectively scored the possible effects of changes in the trait definition should be considered. Overall changes in definition of traits may be important for valuing current cows according to current standards, but it should be done with care, considering its effect on genetic progress, accuracy of BV and trait harmonisation with Interbull member countries. The impact of change in trait definition on DGV accuracy could be more pronounced if the change in trait definition coincides with the subdivision of bulls into reference and validation set as was case in the current study. A gradual decrease in genetic correlations of a trait over time will most likely affect the data of all bulls and will be realised as low h^2 and consequently low DGV accuracy. For traits where there is marked change in trait definition that result in reduced accuracy of prediction of BV alternative genetic evaluation models including multi-variate and random regression models (Tsuruta *et al*. 2004; Haile-Mariam and Pryce 2015) or ignoring old data (Jamrozik and Schaeffer, 1991) should be considered.

In conclusion for most traits s the discrepancy between the adjusted and expected accuracy of DGVs can be related to change in trait definitions. However, there were cases where differences between observed and expected accuracy could not be related to the absence or presence of change in trait definition that can be detected by calculating genetic correlations.

ACKNOWLEDGEMENTS
We thank the ADHIS and Holstein Australia for provision of data used in this study.

REFERENCES

Haile-Mariam, M., G.J. Nieuwhof, K.T. Beard, K.V. Konstantinov, and B.J. Hayes. (2013) *J. Anim. Breed. Genet*. **130**: 20.

Haile-Mariam, M., J.E. Pryce, C. Schrooten, and B.J. Hayes. (2015) *J. Dairy Sci*. **98**:3443.

Haile-Mariam, M., and J.E. Pryce. (2015) *J. Dairy Sci*. Accepted.

Hayes, B.J., H. D. Daetwyler, P. Bowman, G. Moser, B. Tier, R. Crump, M. Khatkar, H. W. Raadsma, and M. E. Goddard. (2009) AAABG *18*:34.

Jamrozik, J. and L. R. Schaeffer. (1991) *J. Dairy Sci.* **74**:1993.

Tsuruta S., I. Misztal, T.J. Lawlor, and L. Klei (2004) *Livest. Prod. Sci*. **91**:199.

Proc. Assoc. Advmt. Breed. Genet. 21: 161-164

PREDICTING GENOMIC SELECTION ACCURACY FROM HETEROGENEOUS SOURCES

J.H.J. van der Werf[1,2], S.A. Clark[1,2] and S.H. Lee[1,2]

[1]School of Environmental and Rural Science, University of New England, Armidale, Australia, 2351
[2]Cooperative Research Centre for Sheep Industry Innovation, Armidale, 2351

SUMMARY

We predict genomic selection accuracy from a heterogeneous reference population that contains close relatives, herd- or flock mates and individuals from the wider population, using an established theory. The various sources of information were modeled as different and independent reference populations with different effective sizes. We show that information on close relatives can have a substantial effect on genomic prediction accuracy. We also show the increase of the genomic prediction accuracy to be less reliant on higher marker density or total reference population size when there are more closely related individuals to predict from. Conversely, the value of close relatives is smaller when the total reference population size is larger. Our modelling is useful to assess the value of a population reference versus a breeder's own reference, based on own animals genotyped.

INTRODUCTION

Genomic selection requires a reference population of individuals having information on both genotype and phenotype. The accuracy of genomic prediction depends on various parameters, including size of the reference sample, its genetic structure and the genetic architecture of the trait of interest. An important parameter is the effective size of the population. The effective population size is a predictor of the effective number of chromosome segments that are represented in the population. Theoretical predictions have usually considered a homogeneous population. However, in most practical applications, the reference population used for genomic predictions possibly consists of many subpopulations, e.g. breeds, lines or strains within a breed and part of the reference population maybe be directly related via pedigree to the animals to be predicted. Hence, reference populations consist of individuals that vary in relatedness to each other and to the target animals to predict. The distinction could be relevant for a breeder with genotyped individuals to assess the importance of own measurement versus that in the wider population.

Clark *et al.* (2012) showed that genomic predictions are more accurate if the genomic relationship between the target animal and the reference population is higher. Habier *et al* (2013) distinguished between three types of information in genomic prediction; linkage disequilibrium, additive-genetic relationships and co-segregation of QTL predicted from marker genotypes within a pedigree. They argued that it would be useful to understand how these sources contribute to the accuracy of genomic predictions, especially when designing reference populations for breeding programs. They show these contributions via simulated examples but did not provide simple predictions for them. Hayes *et al.* (2009) also considered the influence of relationships on genomic prediction. They followed the same approach as the general theory, i.e. by considering the number of independently segregating chromosome segments within families. They showed the accuracy of genomic prediction from varying sizes of full- and half- sib families, but did not consider the information from combined sources. We propose a simple approach to assess the importance of various sources of information used for genomic prediction in animal breeding.

MATERIALS AND METHODS

Predicting genomic selection accuracy. The accuracy of genomic breeding values (GBV) based on DNA marker genotypes can be predicted from theory (e.g. Daetwyler *et al*., 2008; Goddard, 2009; Goddard *et al*., 2011), assuming that prediction is based on a reference population of animals with phenotypes and genotypes for the same DNA markers, and these markers are linked to quantitative trait loci (QTL). Based on the infinitesimal model, the accuracy depends on *i)* the proportion of genetic variance at QTL captured by markers and *ii)* the accuracy of estimating marker effects. The proportion of genetic variance at QTL captured by markers (b) depends on LD between markers and QTL, which in turn depends on the number of markers (M) and the number of 'effective chromosome segments' (M_e); $b = M/(M_e + M)$. Prediction of M_e is not easy and various approximations have been presented by largely the same authors (Goddard, 2009; Hayes *et al.,* 2009, Goddard *et al*., 2011, Meuwissen *et al*., 2013). We will use $M_e = 2N_E Lk/\ln(2N_E)$ (Meuwissen *et al*., 2013), where N_E = effective population size; L = average chromosome length; k = number of chromosomes. The accuracy of estimating marker effects depends on the captured genetic variance as a proportion of the total variance ($b.h^2$), the number of (unrelated) animals observed in the reference population (T), and M_e. The accuracy is the variance of the estimated (random) marker effects (q) as a proportion of the variation in true marker effects: $V(\hat{q})/V(q)$. This term is estimated as $\theta/(1+\theta)$, where $\theta = Th^2b/M_e$. Reliability of GBV is then $r^2 = b.V(\hat{q})/V(q)$ and the accuracy is the square root of this value.

Effective population size in a heterogeneous population. A critical parameter in the accuracy of genomic prediction is the effective population size (N_E). It is not easy to define 'population' in many practical cases and it is not possible to represent a reference population by a single value for N_E. We propose a very simple model relevant for breeding programs for beef cattle or sheep. For the prediction of an individual within a herd/flock we consider three sources of information based on animals that are measured and genotyped *1)* N_1 individuals from a certain breed but not closely related to the target animal*, 2)* N_2 herd/flock mates of the target animal and *3)* N_3 close relatives of the target animal. We will refer to these sources of information as *breed, flock* and *relatives*, respectively. This is, of course, a simplified representation of heterogeneity, but a useful start to consider the contribution of each to overall prediction accuracy. We consider these three subsets as populations that differ in relatedness to the target animal as well as to each other, to be modeled as three different populations with different effective size, indicated as N_{E1}, N_{E2}, and N_{E3}, and a different number of chromosome segments, i.e. differing also in the size of the segments shared amongst each other and with the target animal. Each of these sources provides an estimation of breeding value and the reliability (r^2_i) of each GBV_i can be calculated as above. The three information sources are combined as $GBV = \Sigma GBV_i$ by using $cov(GBV_i, GBV_j) = r^2_i. r^2_j.V_A$, and $cov(GBV_i, a) = r^2_i.V_A$, where a is the true breeding value and V_A is the additive genetic variance. The accuracy of the GBV can then be calculated using standard selection index theory.

Study Design. For each of the three resources contributing to genomic prediction we varied values for N_{Ei} and N_i and marker density. We compared accuracy of GBV from just *breed* with predictions that included also information from *flock* and *relatives*. The total number in the reference population was kept equal between such comparisons. We evaluated the contribution of each information source as 'value of variate', defined as the relative loss in accuracy if that resource was removed. The trait heritability was 0.25.

RESULTS AND DISCUSSION

In a base scenario we assumed a population with a large diversity, N_{E1} = 1000, e.g. similar to the Merino population. Subsets of flock mates and relatives were represented by N_2 = 400 and N_3 = 50, with effective size N_{E2} = 50 and N_{E3} = 8. This scenario represents a lower value for the *breed* information source due to its large diversity, and a large number of individuals in the *flock* and *relatives* information sources. Results are shown in Figure 1, showing that the *flock* and *relatives* resources contribute substantially to the prediction accuracy, especially when the accuracy of the *breed*

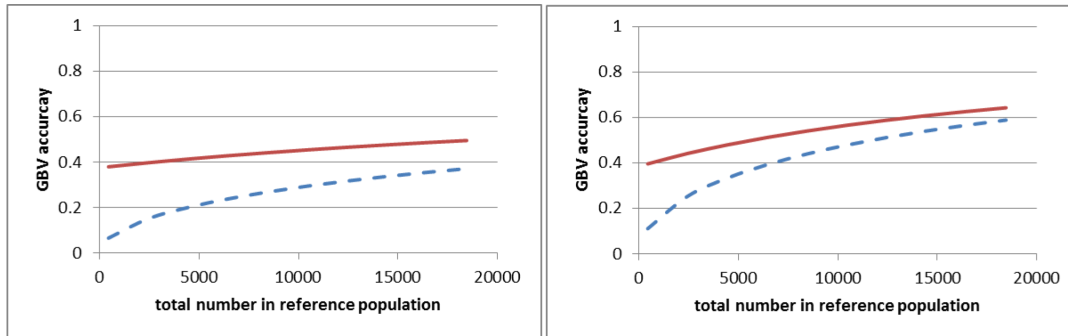


Fig. 1. Accuracy of GBV depending on total reference population size for low (Nmarkers=12k, left) and high (Nmarkers = 500k, right) marker density, comparing 'with' (continuous line) and 'without' (dashed line) information on *flock* and *relatives*.

resource is low. This is the case with low N_1 and with low marker density coupled with high population diversity (N_{E1}). The influence of *flock* and *relatives* decreases with large N_1 and also with higher marker density. Further comparisons are summarized in Table 1. The results show that for populations with lower N_{E1} the contribution of *flock* and *relatives* declines rapidly. If the contribution of *flock* and *relatives* is smaller due to less own data being available (lower values for N_2 and N_3) then their influence decreases accordingly, but it can still be substantial for small N_1 and high N_{E1}.

Overall the results illustrate that the GBV accuracy is likely higher than predicted based on the size in the reference population and the effective population size of the breed, due to information from relatives and more closely related individuals in the flock or herd. The effect will be larger when the information from the wider *breed* resource is of lower value, e.g. for smaller reference populations, or breeds with higher diversity. The effect of marker density is more notable if the breed diversity is high (high N_{E1}). The information from the own flock genotyping and recording can contribute substantially, and even if the numbers are relatively low (low N_2 and N_3) if the *breed* resource is small (e.g. N_1 = 2000). The assumption about N_{E2} and N_{E3} have some effect on the observed differences, e.g. when N_{E3} increases from 8 to 16 in the first case, the accuracy increase (diff) reduced from 95% to 87% and when N_{E2} increases from 50 to 100, the increase is further reduced to 64%.

The purpose of this study was to use a simple model to estimate of the importance of information on closer relatives in genomic prediction. This is relevant for breeders that have developed their own reference population. The value of this own reference can be substantial, unless a fairly large breed reference is available, and the value would be higher for more diverse breeds such a Merino.

Table 1 Value of the various information sources, accuracy of GBV with and without the *flock* and *relatives* information sources[2] and the relative accuracy difference (diff).

N_1	Value of information source[1] *breed*	*flock*	*relatives*	GBV_acc_with	GBV_acc_wo	diff[3]
N_{E1}=1000, N_2=400, N_3=50						
2,000	16%	52%	21%	0.428	0.220	95%
5,000	31%	39%	15%	0.471	0.318	48%
10,000	45%	26%	10%	0.528	0.420	26%
N_{E1}=1000, N_2=100, N_3=10						
2,000	48%	36%	12%	0.279	0.205	36%
5,000	68%	19%	6%	0.357	0.309	15%
10,000	79%	11%	4%	0.445	0.414	7%
N_{E1}=200, N_2=400, N_3=50						
2,000	45%	26%	10%	0.528	0.448	18%
5,000	62%	12%	5%	0.640	0.599	7%
10,000	72%	5%	2%	0.739	0.718	3%

[1] Percent decrease in accuracy if this information source was removed. Note that these do typically not add up to 100%.
[2] N_{E2} = 50, N_{E3} = 8, Marker density = 50k.
[3] Difference between prediction accuracy with and without information from flock and relatives

CONCLUSIONS

This work shows a simple approach for modeling genomic prediction in a heterogeneous reference population by considering several subpopulations that differ in effective size. The model allows quantification of the importance of the own flock or herd information versus the wider breed information used for genomic prediction. We show that as a result of using some information from relatives, the increase of prediction accuracy with increasing the size of the wider reference population, or increasing marker density, maybe lower than expected. The validity of the approach needs to be tested with simulated as well as real data.

ACKNOWLEDGEMENTS

SHL is partly funded by the Australia Research Council (DE130100614)

REFERENCES

Clark, S.A., Hickey, J.M., Daetwyler, H.D. and van der Werf, J.H.J. (2012) *Genet. Sel. Evol*. **44**: 4.
Daetwyler, H.D., Villanueva, B. and Woolliams, J.A. (2008) *PLoS ONE* **3**(10):e3395
Goddard, M.E. (2009) *Genetica* **136**: 245.
Goddard, M.E., Hayes, B.J. and Meuwissen, T.H.E. (2011) *J. Anim. Breed. Genet.* **128**: 409.
Habier, D., Fernando, R.L. and Garrick, D.J. (2013) *Genetics* **194**: 597
Hayes, B.J., Visscher, P.M. and Goddard, M.E. (2009) *Genet. Res*. **91**:47.
Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2013) **Annu. Rev. Anim. Biosci**. **1**: 221.

Proc. Assoc. Advmt. Breed. Genet. 21: 165-168

ACCURACY OF GENOMIC PREDICTION FOR MERINO WOOL TRAITS USING HIGH-DENSITY MARKER GENOTYPES

N. Moghaddar[1, 2], A.A. Swan [1, 3] and J.H.J. van der Werf[1, 2]

[1] Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW 2351, Australia.
[2] School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia.[3] Animal Genetics & Breeding Unit (AGBU), University of New England, Armidale, NSW 2351, Australia

SUMMARY

High-density (HD) marker genotypes could increase the accuracy of genomic prediction by providing stronger linkage disequilibrium (LD) between markers and quantitative trait loci affecting a trait, especially in populations with a high genetic diversity such as Australian Merino sheep. The aim of this study was to compare the accuracy of genomic prediction for Merino yearling and adult wool traits based on observed and imputed 600K single nucleotide polymorphism (SNP) marker genotypes with the accuracy based on moderate-density (50K) marker genotypes. Genomic best linear unbiased prediction (GBLUP) and a Bayesian approach (BayesR) were used as prediction methods. Results showed a small relative increase in accuracy between 2 to 15% (of the previous accuracy) when using a HD marker set. The results of BayesR were on average similar to GBLUP. Considerably higher (up to 25% relative increase) in prediction accuracy was observed for animals with lower genomic relationship to the reference population.

INTRODUCTION

Genomic prediction of selection candidates (Meuwissen, *et al.* 2001) is becoming more practical in animal breeding programs. Genomic prediction is based on genome-wide single nucleotide polymorphism (SNP) marker genotypes assumed in LD with quantitative trait loci (QTLs) affecting a polygenic trait. Genomic prediction based on denser SNP panels is expected to improve the prediction accuracy and hence the selection response compared with using lower-density markers because of a higher LD between markers and QTLs. Higher marker density could be more important in more genetically diverse breeding populations such as Australian Merino sheep, in which the effective population size is reported to be large (Kijas *et al* 2012). The objective of this study is to compare the accuracy of genomic prediction between a HD (600K) and a moderate-density (50K) SNP marker panel for wool traits in Australian Merino sheep using either Genomic Best Linear Unbiased Prediction (GBLUP) or a non-linear Bayesian prediction approach.

MATERIALS AND METHODS

Reference population, phenotypes and validation population. The investigated traits were yearling and adult wool quantity and quality traits as summarized in Table 1. The size of the reference population for each trait and age group was different, ranging from 2,413 to 4,662 purebred Merinos. These animals belonged to the Sheep Cooperative Research Centre Information Nucleus Flock (INF) and the Sheep Genomics Flock (SGF). The INF consisted of eight flocks located across different regions of Australia and these were linked to each other by using common sires through artificial insemination between 2007 and 2011 (van der Werf *et al.* 2010). The SGF was a single research flock located in southern New South Wales, Australia with data collection in 2005 and 2006 (White *et al*, 2011).

The validation population was a group of 175 Merino sires with highly accurate EBVs (average accuracy ~ 0.92). Furthermore, the validation population was divided into two sets of animals; one with a high genomic relationship to the reference population (mean of top 30 relationships was greater than 0.20) and one with a low genomic relationship to the reference population (maximum genomic relationship was less than 0.10).

Genotypes. Genotypes were available based on the 50K Ovine marker panel (Illumina Inc., San Diego, CA, USA). This marker panel provided 48,559 SNP genotypes after applying quality control on genotypes. All INF and SGF sires and a number of progeny (1,735 purebred and crossbred Merino animals) were genotyped using the 600K (Illumina Inc., San Diego, CA, USA) marker panel, which provided 510,174 SNPs after quality control. Using animals with observed HD genotypes as an imputation reference set, the rest of Merinos were imputed from 50K to 600K using FImpute (Sargolzaei 2014).

Statistical methods. Genomic best linear unbiased prediction (GBLUP) and a BayesR approach (Erbe *et al.* 2012) were used to calculate the Genomic Breeding Values (GBV) using ASReml (Gilmour *et al.* 2009) and BESSiE (Boerner and Tier, 2015), respectively. The following model was used for data analysis: $y = Xb + Z_1g + Z_2m + Z_1Qq+e$ where y is a vector of phenotypes, b is a vector with fixed effects, g is the random additive genetic effect of the animal, m is a vector with maternal effects, q is a vector of genetic groups and e is vector of random residual effects, X, Z_1 and Z_2 are incidence matrices. g, m, q and e are considered normally distributed as $g \sim N(0, G\sigma_g^2)$, $m \sim N(0, I\sigma_m^2)$, $q \sim N(0, I\sigma_q^2)$ and $e \sim N(0, I\sigma_e^2)$, respectively, where G is the genomic relationship matrix calculated based on 50K or 600K genotypes using the VanRaden (2008) approach. The fixed effects in the model were birth type, rearing type, gender, age at measurement (for weaning weight and post weaning weight) and contemporary group which was flock × birth year × management group.

Table 1. Summary statistics and heritability of yearling (Y) and adult (A) wool traits.

Trait	Nr. records	Mean	s.d	Range	$*h^2$
Y[1]-GFW	4,662	3.64	1.04	1.2 - 7.8	0.57 (0.04)
Y-CFW	4,423	2.46	0.65	0.93 - 4.76	0.51 (0.05)
Y-FD	3,969	19.93	5.39	12.8 - 42	0.62 (0.04)
Y-FDCV	3,554	19.26	2.86	11.7 - 31.8	0.47 (0.04)
Y-SS	3,554	33.8	9.82	13 - 88	0.55 (0.04)
Y-SL	3,554	80.93	13.06	38 -236	0.56 (0.04)
A[2]-GFW	4,541	5.75	1.97	1.50 - 14.30	0.69 (0.04)
A-CFW	4,540	4.19	1.39	1.13 - 9.91	0.70 (0.04)
A-FD	3,001	18.17	1.84	13.80 - 24.60	0.64 (0.05)
A-FDCV	2,436	18.07	2.56	11.80 - 27.70	0.57 (0.07)
A-SS	2,414	36.61	10.31	3.00 - 68.00	0.37 (0.07)
A-SL	2,413	98.57	18.34	41.00 - 149.00	0.67 (0.07)

[1]Y=Yearling, [2]A=Adult, GFW=Greasy Fleece Weight$_{(Kg)}$, CFW=Clean Fleece Weight$_{(Kg)}$, FD=Fibre Diameter$_{(\mu)}$, FDCV=Fibre Diameter Coefficient of Variation $_{(\%)}$, SS=Staple Strength$_{(Newton/ktex)}$, SL= Staple Length$_{(mm)}$,*: estimated based on pedigree

The BayesR method considers a mixture of four normal distributions for the SNP effects with variances $\sigma_1^2 = 0, \sigma_2^2 = 0.0001\,\sigma_g^2$, $\sigma_3^2 = 0.001\sigma_g^2$, $\sigma_4^2 = 0.01\sigma_g^2$. Starting values for σ_g^2 were taken from GBLUP analysis and the priors for the proportion of markers in each distribution was drawn from a Dirichlet distribution. 50,000 iterations (with 10,000 burn-in) were run for analysis. The

genomic prediction accuracy was assessed based on the Pearson correlation coefficient between GBV of the validation sires and their accurate EBV based on progeny test.

RESULTS AND DISCUSSION

The accuracy of genomic prediction for the two marker panel densities is shown in Tables 2 and 3 for yearling and adult wool traits, respectively, based on GBLUP and BayesR prediction methods. Results showed a slight increase in accuracy for both yearling and adult wool traits based on HD genotypes. The relative increase in prediction accuracy was ranging from 2% to 15% with an average relative increase of 5.9%. The percentage point of gain in accuracy was between 0.00 and 0.09 and on average 0.04. BayesR did not show notably higher accuracies than GBLUP based on 600K across all yearling and adult wool traits.

Table 4 shows the change in GBV accuracy for groups of validation sires with high or low genetic relationship to the reference population. A considerable increase in accuracy was observed across almost all traits for animals with lower genetic relationship to the reference population, while the increase in accuracy for highly related animals was small.

This study showed a small gain in GBV accuracy based on HD genotypes in Merino sheep, except for animals with lower genetic relatedness to the reference population in which extra accuracy was notable. As Table 3 and 4 show, the genomic prediction of wool traits based a moderate-density marker set (50K) is already high (up to 0.68) which is because of a relatively high genetic relatedness of validation sires to the reference population. This indicates for highly related animals a moderate density marker panel (~50K) could explain most of the additive genetic variance of the wool traits used in this study.

Results showed significantly higher GBV accuracy based on HD genotypes for lowly related animals to reference population. Animals with lower relatedness share smaller chromosome segments and rely more on higher marker density to achieve sufficient LD for accurate genomic prediction.

Table 2. Accuracy of genomic prediction based on using 50K or 600K marker genotypes in yearling wool traits.

Trait	Size	GBV Accuracy		
		GBLUP (50k)	GBLUP(600k)	Bayes-R(600k)
Y[1]-GFW	4,662	0.681	0.692	0.669
Y-CFW	4,423	0.621	0.634	0.632
Y-FD	3,969	0.686	0.752	0.718
Y-FDCV	3,554	0.462	0.469	0.470
Y-SS	3,554	0.366	0.412	0.369
Y-SL	3,554	0.594	0.617	0.621

[1]Y=Yearling, GFW=Greasy Fleece Weight$_{(Kg)}$, CFW=Clean Fleece Weight$_{(Kg)}$, FD=Fibre Diameter$_{(\mu)}$, FDCV=Fibre Diameter Coefficient of Variation $_{(\%)}$, SS=Staple Strength$_{(Newton/ktex)}$, SL= Staple Length$_{(mm)}$

Genotype imputation errors might be a potential reason of limiting gain in GBV accuracy from HD genotypes. However the chance of this error should be very low in this study because the HD genotyped animals (1,735) were selected based on high genetic relationships to the rest of population. Furthermore, our previous results showed high imputation accuracy of low-density (12K) to moderate density (50K) genotype if there is a high genetic relatedness between test set and imputation reference set (Moghaddar *et al.* 2015). Imputation of a moderate (50K) to high density (600K) is expected to be more accurate than imputation of low to moderate marker density.

Table 3. Accuracy of genomic prediction based on using 50K or 600K marker genotypes in adult wool traits.

Trait	Size	GBV Accuracy		
		GBLUP (50K)	GBLUP(600K)	Bayes-R(600K)
A[1]-GFW	4,541	0.650	0.691	0.691
A-CFW	4,540	0.594	0.631	0.626
A-FD	3,001	0.610	0.673	0.703
A-FDCV	2,436	0.324	0.366	0.370
A-SS	2,414	0.590	0.669	0.664
A-SL	2,413	0.400	0.461	0.464

[1]A=Adult, GFW=Greasy Fleece Weight(Kg), CFW=Clean Fleece Weight(Kg), FD=Fibre Diameter(µ), FDCV=Fibre Diameter Coefficient of Variation (%), SS=Staple Strength(Newton/ktex), SL= Staple Length(mm)

Table 4. GBV accuracy for genetically highly or lowly related animals to reference population

Trait	50K-Marker Density		600K-Marker Density	
	Highly Related	Lowly Related	Highly Related	Lowly Related
Y-GFW	0.712	0.398	0.721	0.410
Y-FD	0.667	0.665	0.766	0.754
Y-SS	0.471	0.226	0.496	0.261
Y-SL	0.720	0.190	0.733	0.237
A-GFW	0.712	0.512	0.712	0.608
A-FD	0.690	0.570	0.735	0.628
A-SS	0.760	0.548	0.762	0.617
A-SL	0.573	0.361	0.586	0.452

GFW=Greasy Fleece Weight$_{(Kg)}$, CFW=Clean Fleece Weight$_{(Kg)}$, FD=Fibre Diameter$_{(\mu)}$, FDCV=Fibre Diameter Coefficient of Variation $_{(\%)}$, SS=Staple Strength$_{(Newton/ktex)}$, SL= Staple Length$_{(mm)}$

ACKNOWLEDGEMENTS

The authors are thankful to K. Gore (AGBU) for database management and performing quality control, to H. Daetwyler and B. Sunduimijid (ECODEV, VIC) for HD genotype imputation, to the project executive staffs at research flocks and to Australian Wool Innovation (AWI) and Meat and Livestock Australia (MLA) who made the data available from the Sheep Genomic Project.

REFERENCES

Boerner V., Tier B. (2015) *AAABG-2015 submitted.*

Erbe M., Hayes B.J., Matukumalli L.K., *et al* (2012) *J. Dairy Sci.* **95**: 4114.

Gilmour A.R., Gogel B.J., Cullis B.R., *et al* (2009) ASReml User Guide Release 3.0

Kijas J.W., Lenstra J.A., Hayes B.J., *et al* (2012) PLoS Biol.10:e1001258

Meuwissen T.H.E., Hayes B.J., Goddard M.E. *(2001) Genetics **157**, 1819*

Moghaddar N., Gore K.P., Daetwyler H.D., Hayes B.J., van der Werf J.H.J. (2015). *Genet.Sel.Evol* (*submitted)*

Sargolzaei M., Chesnias J.P., Shenkel F.S. (2014) *Genomics* 2014, **15**:478

Van der werf J.H.J., Kinghorn B.P, Banks R.G. *(2010) Animal Production Science* **50**, 998

VanRaden P.M. (2008) *J. Dairy Sci* **91:** 4414.

White J.D, Allingham P.G, Gorman C.M., *et al* (2012) *Anim.Prod.Sci* **52**:157

Proc. Assoc. Advmt. Breed. Genet. 21: 169-173

CONSIDERATIONS IN THE USE OF COMMERCIAL DATA FOR GENETIC EVALULATION IN BEEF AND SHEEP IN AUSTRALIA

R. G. Banks and B. Tier

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

SUMMARY

Using data collected on commercial animals for genetic evaluation is appealing due to the larger volumes of data potentially available, and the fact that such data may be on objective traits, or traits more closely correlated with consumer end-point traits. However, there are important considerations, the most significant being the degree of genetic relationship between commercial animals and current candidates for selection. There will continue to be advantages in using data collected in structured programs, rather than relying on commercial data.

INTRODUCTION

There is growing interest in the use of commercial data in genetic evaluation of beef cattle and sheep, stimulated in part by the proliferation of means of capturing and transporting data and of interrogating databases at lower cost, coupled with broader discussion of "big data". In general terms, big data refers to finding patterns or associations between factors in large datasets, with a lively debate continuing over what potential this offers and which traditional approaches to finding meaning will be rendered redundant.

One application is in livestock production, where databases built for different purposes are starting to be linked through the common identity provided by NLIS, (National Livestock Identification Scheme: - for example data collected in processing plants for meat eating quality assessment (Meat Standards Australia, or MSA)), and with genetic databases such as those containing pedigree, performance and genotype information for use in BREEDPLAN and Sheep Genetics evaluations. The appeal behind this area of application has several elements:

- Wanting to get the most out of the substantial investments in building these data systems
- Seeking to engage commercial producers more directly with genetic information, in part by providing a means for them to contribute to the evaluations, and
- Seeking to exploit data that is already captured on otherwise hard-to-measure (HTM) traits related to carcase and meat characteristics, and potentially animal health data.

This paper briefly explores this opportunity and highlights important considerations that impact the value of the enterprise of increasing connections between various data systems, to generate big data.

For this exploration, it is important to briefly summarise the genetic structure of the beef and sheep industry populations, but focussing mainly on beef.

INDUSTRY GENETIC STRUCTURE

In both the beef cattle and sheep industries, there are well-defined and separate sire breeding, or seedstock, and commercial production sectors, with much larger numbers of animals in the commercial sector. For example, the national commercial cow herd is approximately 12m head and includes approximately 100,000 enterprises. The seedstock sector comprises approximately 250-300,000 animals in 1,500-2,000 enterprises.

The sires of commercial animals – herd bulls in the beef industry (and flock rams in sheep) –

---

* AGBU is a joint venture of NSW Department of Primary Industries and University of New England

are in simple terms bulls born in seedstock herds that are not retained as sires of sires, and which are marketable. These herd bulls will be bred to approximately 100 cows in a working life. Genetic evaluation is based predominantly on records of performance taken in seedstock herds on pedigree-recorded young bulls and heifers, and similarly in sheep.

Given that the majority of males born will be used as herd bulls – to sire commercial progeny – the closest commercial relatives of young candidate males will be the progeny of half-brothers. If any of these are themselves used to breed herd bulls, this adds 2 more steps in the relationship path between young candidates and commercial relatives. Because AI is essentially only used at the seedstock level, cases where young male candidates have commercial half-sibs that can be recorded are very rare in "normal" commercial practice.

This structure contrasts with that in dairy cattle, where commercial cows being herd-recorded are half-sisters of young bulls, and in pigs, where full- and half-sibs of young candidates can be recorded for slaughter/carcase traits. What does this mean for the value of commercial data in beef cattle and sheep?

ACCURACY OF ESTIMATED BREEDING VALUES USING COMMERCIAL DATA

It is straightforward to calculate potential accuracies of EBVs for various data combinations. Table 1 shows the results for a sample of scenarios.

Table 1. EBV accuracy for candidate animals, with varying heritability and combinations of effective records on relatives.

Heritability of trait	Records available	Accuracy of EBV for candidate animals in seedstock herd/flock
0.4	Own record	0.63
0.4	29,850 grand-progeny of candidate's paternal grand-sire (PGS). PGS has 100 sons each with 30 progeny	0.25
0.4	930 progeny of 1 son of the same paternal grand-sire as the candidate	0.125
0.3	Own record	0.55
0.3	Own record plus 3 half-sibs plus 3 animals with same paternal grand-sire, each with 30 progeny	0.60
0.3	60 progeny of 2 animals that share the same paternal grand-sire	0.15

The overall pattern is simple and clear:

- Information from animals that share the same grand-parents is of limited value for genetic evaluation of young candidates in the seedstock sector
- Recorded and/or progeny-tested half-sibs can add accuracy

This simple example is for the situation where the trait recorded is the same for all animals. Data from relatives is potentially more useful when data can be collected on a trait more closely correlated with an objective trait, or that cannot be recorded on the candidate at all. This situation applies for carcase marbling (or other eating quality traits). Animals in seedstock herds can be indirectly assessed for marbling using live scanning or other correlated traits, but not for the objective trait itself.

In this situation, direct carcase measures on very small numbers of progeny generate more accuracy than even very large numbers of recorded half-sibs, for example. However, the contribution to accuracy of EBV on the candidates depends on the genetic relationship with the

animals recorded for the objective trait – the value of the data declines as the square of the number of steps in the relationship path.

MAKING COMMERCIAL DATA WORTH HAVING

From the perspective of genetic evaluation of young candidate animals in seedstock herds/flocks, the most likely situation to be practical is to collect data on commercial progeny of half-sibs, or on half-sibs themselves. The former requires that each crop of young sires generates commercial progeny which are recorded, the latter that seedstock sires are mated to produce both seedstock and commercial progeny routinely.

Examples of these 2 scenarios exist in beef cattle in Australia:

1. In the Team Te Mania program commercial herds use semen from stud sires of sires, or current young bulls, and capture slaughter data. Accordingly, young bulls are evaluated with data from animals with either the same sire (relationship = 0.5) or grand-sire (0.25)
2. In the Wagyu breed to date, a high proportion of commercial animals are AI progeny of widely used sires. This means that young bulls have commercial half-sibs with data. Zhang (2015, in press) details the data currently available in this breed for genetic evaluation, but in simple terms it is much more like dairy data than beef in terms of the relationship x data pattern.

In both these cases, higher accuracy of genetic evaluation is achieved for breeding objective traits of young seedstock animals than is usual in beef cattle in this country.

The reference populations (or information nucleus herds/flocks) established in beef and sheep combine some aspects of these 2 examples, but with potentially wider reach. In each, elite young sires are being progeny tested including for direct objective traits. The impact on accuracy of young animals in the seedstock population at large then depends on the animals' relationship with the animals being progeny tested. Here the intention is more to generate reference data for genomic selection, which partly overcomes the variable impact on accuracy of evaluation caused by the variation in relationships. At the same time, animals that are progeny-tested in this way will inevitably achieve higher accuracy of EBV for the traits recorded which in turn will increase their likelihood of being selected as parents in the seedstock sector.

CAVEATS

There are two obvious concerns regarding use of commercial data:

1. unless data is collected in identifiable management groups, the heritability of the data will be compromised, and may in fact reach zero
2. such groups must have more than one sire represented, or if genomic pedigree is used, be sufficiently diverse to support some statistical contrast.

These aspects of data have been examined in the Wagyu case, which is to date simply a semi-random sample of commercial datasets, analysed by Zhang *et al.* (2015, pers. comm.). Within the data:

- 5,270 recorded were recorded in 1,161 management groups) for an average group size of 4.5 animals
- 692 management groups (60% of groups) contained 0 effective progeny, but 1,197 recorded animals (or 23% of the data)
- The overall average effectiveness of data (the ratio of effective number to actual) was 24%, with the value for groups with at least 1 effective progeny being 60%.

In general, utility of commercial data will be maximised when:

- Management groups are accurately recorded, and effective and actual progeny numbers are as similar as possible
- The commercial animals are as closely related to selection candidates as possible

- The animals are recorded for objective traits, or traits highly correlated with objectives.

IMPLICATIONS FOR USE OF COMMERCIAL DATA

Even in the Wagyu situation, at least currently inherently favourable for use of commercial data, 60% of data collected (as measured by number of groups or datasets) provides no information that can be used for genetic evaluation.

More generally in the beef and sheep industries, it is not easy to predict the distribution of management group effective size, but there is no obvious reason to expect it to be dramatically different from in this example, and just as importantly, the genetic relationship between records and current candidates is likely to be lower. This last reflects the fact that widespread AI is very unlikely to be commercially practical – certainly use of AI in bull or ram multiplication makes growing sense as high merit, high accuracy sires are identified, but bulls and rams are extremely efficient AI technicians for extensive operations.

If these surmises are correct, it will be important to proceed carefully in harvesting commercial data, or more precisely, in what benefits are promoted from that harvesting. Even if the capture were free, costs will be generated in data storage and analysis. If the capture is not free, it will be imperative to develop, and communicate very clearly, ways of valuing data in advance of its collection so that informed investment decisions can be made.

The discussion to this point presupposes that the only purpose of capturing commercial data is for genetic evaluation. This may not be the case – management decision tools may be developed around real-time commercial data, essentially akin to herd recording in dairy cattle. In this scenario, as long as simple ways of screening in the data that is useful for genetic evaluation can be applied, some benefit is possible.

More generally, the multiplication and dissemination structure of the beef and sheep breeding and production industries inherently favour development of structured data collection. Harvesting data on close relatives of current selection candidates is genetically and hence economically more efficient, with the caveat that the "commercial" conditions must be commercial, otherwise there is a GxE to contend with and possible loss of confidence in the EBVs.

Livestock genetic improvement has been a "quite big data" enterprise ever since the introduction of BLUP methods, and is automatically becoming genuinely big as volumes of genotypic information grow. Because the field has always been focussed on extracting maximum value from precious (ie expensive) data, basic principles for valuing data and for designing efficient recording structures are well developed.

CONCLUSIONS

Increasing use of large volumes of commercial data is very appealing in principle, and becoming more feasible through developments in data transfer and storage, and greater willingness to link databases.

However, "data ain't data" – commercial data will vary enormously in its value for genetic evaluation, and simply assuming that incorporating large amounts of commercial data will lead to dramatic increases in genetic progress is misguided. It is almost certain that the greatest value will come from carefully structured and managed data collection, and those breeding enterprises that can incorporate such activity into their business plans will always be at an advantage.

ACKNOWLEDGMENTS

This research is partly supported by Meat and Livestock Australia.

REFERENCES

Zhang, Y., Tier, B. and Banks, R.G. (2015) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21:** (in press)

Proc. Assoc. Advmt. Breed. Genet. 21: 173-176

WHAT CAUSES THE NEGATIVE GENETIC CORRELATION BETWEEN MATERNAL AND DIRECT EFFECTS FOR LAMB BIRTH WEIGHT?

H.T. Blair
International Sheep Research Centre, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand.

SUMMARY

An antagonistic genetic relationship between direct and maternal effects for birth weight (and weaning weight) makes the concurrent genetic improvement of the two traits more challenging than if the correlation was zero or favourable. The direction and magnitude of the genetic correlation between maternal and direct effects for lamb birth weight are equivocal with several moderate negative and several near zero estimates being reported. A number of experiments over the last decade undertaken at Massey University in the pursuit of developmental programming effects in sheep, have provided serendipitous results that suggest a negative phenotypic correlation between dam and offspring birth weight can be induced by some environmental stressors during pregnancy. These correlations could be interpreted as having an underlying genetic effect, since the relationship is generated without any experimental manipulation of second generation lambs. It is proposed that mining existing datasets might be a fertile ground to conduct research for further exploration of this possible explanation of the negative genetic correlation between direct – maternal effects for birth weight.

INTRODUCTION

The dam – fetal conflict has long been recognised in both domestic animals and humans. A series of papers in the Journal of Animal Science in 1972 reviewed maternal and direct genetic effects in various species and also reviewed techniques for estimating the effects; Bradford (1972) is particularly relevant to this paper. Haig (1993) published a seminal paper entitled “genetic conflicts in human pregnancy” which advanced thinking about the biology of human pregnancy. Haig’s contribution has become encapsulated in the phrase “Haig’s conflict”, which describes the need for dams to control the resources that a foetus demands from its mother. At a similar time Hales and Barker (1992) proposed the thrifty phenotype (also known as the Barker hypothesis) to explain a putative effect of events during pregnancy on new-born and adult phenotypes.

An important feature of the dam – fetal relationship is that many estimates of maternal – direct correlations for both birth weight (and weaning weight – not discussed further) are negative (Table 1; see also reviews by Gootwine *et al*. 2007 and Brien *et al*. 2014). This is suggestive of an evolutionary effect that may avoid risk to the dam of gestating a large foetus that cannot easily pass through the birth canal and/or minimises the opportunity for a species to outgrow its ecological niche by continuing to increase in body size over time. Female mammals with low fecundity must successfully reproduce at least 3 times to maintain population size and therefore they cannot afford to invest all their bodily resources in the current foetus. However, the neonate is often born into a dangerous environment and they want to be large and healthy with energy reserves and therefore it has a drive to scavenge resources.

Since the late 1990s, there has been a growing interest in developmental programming and the likely epigenetic mechanisms (Langley-Evans 2006). While the various epigenetic mechanisms (methylation, acetylation, small RNA’s) are now accepted, there is still much to learn about their roles in intra- and inter-generational effects on animal phenotypes. The phenotypic effects of maternal and paternal imprinting (involving epigenetic mechanisms) have also been recognised in production animals, and it is possible that imprinting contributes to quantitative variation in production traits (Wolf *et al.* 2008).

The purpose of this paper is to report on some serendipitous results for lamb birth weight obtained from a series of experiments designed to identify possible developmental programming effects. One possible explanation of the results is an epigenetic effect for either maternal or direct effects on birth weight.

METHODS

Sheep experiments were undertaken at Massey University from 2005 to 2014 in which birth weights were collected for 2 generations. The first generation (G1) were born to dams which were either exposed to differential feeding during pregnancy (Kenyon *et al.* 2011, Paten *et al.* 2011) or whose dams differed in age (Loureiro *et al.* 2012). Female offspring (G2) were retained, treated as one group and the birth weight of their offspring recorded.

G1 birth weight data were analysed using PROC GLM in SAS with a linear model that included the effects of dam treatment (level of pregnancy feeding or dam age), lamb sex and lamb birth rank and a covariate for date of birth. G2 birth weights were analysed in a similar manner except the treatment effect was that of their grand-dams.

Table 1. Heritability estimates for direct lamb birth weight, maternal effect and the genetic correlation between maternal and direct effects

Author	Maternal	Direct	Correlation
Tosh and Kemp (1994)	0.13 to 0.31	0.07 to 0.39	-0.13 to -0.56
Nasholm and Danell (1996)	0.30	0.07	+0.11
Larsgard and Olesen (1998)	0.42	0.22	-0.10
Yazdi *et al.* (1999)	0.14	0.15	+0.10
Ligda *et al*. (2000)	0.19	0.18	-0.44
Al-Shorepy (2001)	0.33	0.42	-0.60
El Fadili and Leroy (2001)	0.28	0.01	+0.01
Ekiz *et al*. (2004)	0.10 to 0.27	0.09 to 0.33	-0.48 to -0.55
Oliveira Lôbo *et al*. (2009)	0.18	0.42	-0.47
Prince *et al*. (2010)	0.08 to 0.34	0.14 to 0.28	-0.48 to -0.57

RESULTS AND DISCUSSION

In the first experiment where ewes were differentially fed during pregnancy (Kenyon *et al.* 2011), there was an interaction between feeding treatment and birth rank for birth weight such that only G1 twins were affected, with those from restricted-fed dams being lighter (Table 2). In all 4 years of G2 birth weight data (ewes aged 2 to 5 year-old), ewes from restricted-fed dams gave birth to heavier lambs than the ewes from a*d lib*-fed dams. In the second differential feeding experiment (Paten *et al.* 2011), restricted feeding during pregnancy did not decrease G1 birth weight and there were no effects of grand-dam feeding on G2 birth weights (data not shown).

It was unsurprising that the G1 birth weights of lambs born to one-year-old ewes were substantially lighter than lambs born to mature dams (Loureiro *et al.* 2012). Indeed, twin lambs born to one-year-old dams were on average 1.5kg lighter than singleton lambs born to mature dams (3.4±0.14kg versus 4.9±0.18kg). This live weight handicap persisted into maturity with G1 1½ year-old ewes being nearly 10kg lighter than their singleton born counterparts (55.9±1.01kg versus 65.2±1.30kg). However, what was surprising was that these lighter G1 ewes gave birth to heavier G2 lambs in their first three lambings (Table 2).

The above results suggest that the birth weights of lambs born to the stressed dams are often lighter than those from non-stressed dams. However, exceptions did occur, whereby singletons

born in the first differential pregnancy feeding experiment did not have modified birth weight according to their dams pregnancy feeding and also in the second differential feeding trial where differences in birth weight did not eventuate. Similar variable results are commonplace in the literature.

In second generation animals, there is a surprising effect whereby ewes which were born small due to a stress on their dam during pregnancy, themselves go on to give birth to heavier lambs. There are few reports from similar experiments in the literature. Contrary to the above observation, Gardiner *et al.* (2007) reported that lamb birth weight increased by 149g for every 1kg increase in the dam's own birth weight. However, their study did not entail any systematic manipulation of the pregnant dam. Furthermore, analyses considering singleton versus twin born ewes do not show that twin-born ewes produce heavier lambs. These inconsistent outcomes suggest there is some, as yet unexplained, biological phenomenon that occasionally results in the reversal of light birth weight between generations. The challenge is to understand the biological mechanisms that underpin the dam-foetus relationship so that it might be manipulated.

Based on the results presented here it might be worthwhile for those analysing the association between direct and maternal effects to reconsider how the relationship is construed both in biological and biometrical terms. A typical analysis assumes only genetic and environmental effects, whereas it is possible there could be epigenetic effects acting on birth weight via either the direct growth genes or the maternal uterine genes. Geoghegan and Spencer (2013) proposed a simple model that could be developed to examine epigenetic effects, while Goddard and Whitelaw (2014) suggested that it might not be necessary to change the way in which animal genetic merit is predicted in the presence of epigenetic effects. However, when there are two traits in a selection objective that are apparently antagonistically genetically correlated, genetic gain in each of those traits will be less than when compared to a situation where the traits are favourably correlated. The question then arises as to whether a seemingly antagonistic genetic correlation that is caused by an epigenetic effect can either be accounted for or ignored (r_g set to 0).

Table 2. Lamb birth weights of dams which were either differentially feed during pregnancy or were of different ages (G1) and the birth weights of their offspring (G2)

Dam treatment	G1 birth weight (kg)	G2 birth weight (kg)			
Pregnancy feeding		2007	2008	2009	2010
Ad lib	6.0 (single)				
	5.1 (twin)	4.3	4.5	4.8	5.0
Restricted	5.9 (single)				
	4.6 (twin)	4.7	4.8	5.2	5.2
Std error	0.07	0.10	0.10	0.11	0.11
Dam age		2011	2012	2013	
Mature	4.6	5.1	5.1	5.7	
One-year-old	3.7	5.6	5.7	6.0	
Std error	0.15	0.11	0.09	0.13	

Various commentators have noted that much research effort is being devoted to exploring epigenetic effects. However, repeatable farm animal paradigms are sparse in the literature, experimental costs are significant and experiments take years to complete (Kenyon and Blair 2014). As a first step, it might be worthwhile to mine datasets used to estimate genetic parameters, although in the absence of a systematic manipulation of the dam during pregnancy to induce a

significant change in birth weight, it may be difficult to detect swings in birth weight such as those reported in Table 2. It would seem worthwhile for those with quantitative animal breeding and genetics skills to brainstorm with those having interests in epigenetic mechanisms to examine whether current models of some quantitatively inherited traits need to be reconfigured. If the biological mechanisms underlying the apparent negative genetic correlation between maternal and direct effects for birth weight (and weaning weight) can be untangled it may be possible to directly manipulate the mechanisms to benefit animal growth and consequently farm profit.

ACKNOWLEDGEMENTS

The efforts of colleagues and postgraduate students in collecting data are gratefully acknowledged. The projects were supported by funding from the Massey University's Sheep Research Centre, Gravida (a Centre for Research Excellence) and Beef and Lamb New Zealand.

REFERENCES

Al-Shorepy S.A. (2001) *Small Ruminant Res.* **39**:219.
Bradford G.E. (1972) *J. Anim. Sci.* **18**:1324.
Brien F.D., Cloete S.W.P., Fogarty N.M., Greeff J.C., Hebart M.L., Hiendleder S., Edwards J.E.H., Kelly J.M., Kind K.L., Kleemann D.O., Plush K.L. and Miller D.R. (2014) *Anim. Prod. Sci.* **54**:667.
Ekiz B., Özcan M. and Yilmaz A. (2004) *Turk. J. Vet. Anim. Sci.* **28**:383.
El Fadili M. and Leroy P.L. (2001) *J. Anim. Breed. Genet.* **118**:341.
Gardner D.S., Buttery P.J., Daniel Z. and Symonds M.E. (2007) *Reproduction* **133**:297.
Geoghegan J.L. and Spencer H.G. (2013) *Theor. Popul. Biol* **88**:1.
Goddard M.E. and Whitelaw E. (2014) *Front. Genet.*, | doi: 10.3389/fgene.**2014**.00247
Gootwine E., Spencer T.E. and Bazer F.W. (2007) *Animal* **1**:547.
Haig D. (1993) *Q. Rev. Biol.* **68**:495.
Hales C.N. and Barker D.J. (1992) *Diabetologia* **35**:595–601.
Langley-Evans S.C. (2006) *Proc. Nut. Soc.* **65**:97.
Larsgard A.G. and Olesen I. (1998) *Livest. Prod. Sci.* **55:**273.
Ligda Ch., Gabriilidis G., Papadopoulos Th. and Georgoudis A. (2000) *Livest. Prod. Sci.* **67**:75.
Loureiro M.F.P., Pain S.J., Kenyon P.R., Peterson S.W. and Blair H.T. (2012) *Anim. Prod. Sci.* **52**:552.
Kenyon P.R., van der Linden D.S., Blair H.T., Morris S.T., Jenkinson C.M.C., Peterson S.W., Mackenzie D.D.S. and Firth E.C. (2011) *Small Ruminant Res.* **97**:21.
Kenyon P.R. and Blair H.T. (2014) *Small Ruminant Res.* **118**:16.
Näsholm A. and Danell O. (1996) *J. Anim. Sci.* **74**:329.
Oliveira Lôbo A.M.B., Braga Lôbo R.N, Rezende Paiva S., Pinheiro de Oliveira S.M. and Facó, O. (2009) *Genet. Mol. Biol* **32**:761.
Paten A.M., Asmad K., Loureiro M.F.P., Kenyon P.R., Pain S.J., Peterson S.W., Pomroy W.E., Scott I. and Blair H.T. (2011) *Proc.NZ Soc. Anim. Prod.* **71**:56.
Prince L.L.L., Gowane G.R., Chopra A. and Arora A.L. (2010) *Trop. Anim. Health Pro.* **42**:1093.
Tosh J.J. and Kemp R.A. (1994) *J. Anim. Sci.* **72**:1184.
Wolf J.B., Cheverud J.M., Roseman C. and Hager R. (2008) *PLoS Genet* **4(**6): e1000091. doi:10.1371/journal.pgen.1000091
Yazdi M.H., Johansson K., Gates P., Nasholm A., Jorjani H. and Liljedahl L.E. (1999) *J. Anim. Sci.* **77**:533.

Proc. Assoc. Advmt. Breed. Genet. 21: 177-180

GENETIC CORRELATIONS BETWEEN PUREBRED AND CROSSBRED PERFORMANCE OF POLL DORSET SHEEP

D.J. Brown[1,3] and J.H.J. van der Werf[2,3]

[1]Animal Genetics and Breeding Unit*, University of New England Armidale NSW 2351
[2]School of Environmental and Rural Science, University of New England Armidale NSW 2351
[3]Cooperative Research Centre for Sheep Industry Innovation Armidale NSW 2351

SUMMARY

The genetic correlation between purebred and crossbred performance has implications for the utility of Australian Sheep Breeding Values (ASBVs) and the accuracy of genomic predications. The aim of this study was to estimate the genetic correlations between purebred and crossbred performance for terminal sire breeds. The genetic correlations estimated between purebred and crossbred progeny performance were all very high ranging from 0.44 to 1.00 and not significantly different from one for the weight and carcase traits examined in this study. These result support the use of the LAMBPLAN across breed ASBVs produced from animals with variable breed composition and also the use of crossbred animals in the genomic reference populations.

INTRODUCTION

Historically the LAMBPLAN genetic evaluation was based mostly on performance data from purebred flocks. The composition of ram breeding flocks is slowly changing with recent drops of rams becoming more composite in their breed makeup. In the 2014 drop of animals in the LAMBPLAN terminal sire analysis, only 24% of the animals recorded were purebred based on the pedigree information supplied.

ASBVs are used by commercial ram buyers across a wide variety of production systems but most likely these will be with the aim to produce crossbred progeny of varying breed composition. Thus it is important that the ASBVs predict sire ranking both in purebred and crossbred progeny. Ingham *et al.* (2005) demonstrated that LAMBPLAN ASBVs were moderately to highly correlated with crossbred progeny performance in maternal breeds. Banks *et al.* (2009) using data from terminal sire breeds observed similar correlations, however they highlighted the large variation in outcomes across traits and sites. Wei *et al.* (1991) pointed out that the correlation between purebred and crossbred performance (r_{PC}) depends on the amount of dominance in a trait, the distance between breeds (differences in allele frequency) and is also often confounded with genotype by environment interaction (GxE).

Furthermore most of the reference populations that underpin the Australian genomic evaluations for terminal sires are based on crossbred progeny (van der Werf *et al*. 2010). Thus an r_{PC} value of less than 1.0 could mean that the genomic breeding values derived from crossbred performance and ASBVs based on purebred performance could have different accuracies depending on what the breeding goal traits are. The aim of this study was to estimate the genetic correlations between the performance of purebred terminal sires and the performance of their crossbred progeny.

MATERIALS AND METHODS

Data. Pedigree and performance data were extracted from the Sheep Genetics LAMBPLAN terminal sire database (SG) (Brown *et al.* 2007). This database consists of pedigree and performance records submitted by Australian terminal sire ram breeders, and is used for genetic

---

* AGBU is a joint venture of NSW Department of Primary Industries and University of New England

evaluation purposes. The database also contains information from the Sheep CRC Information Nucleus Flocks (INF) where all progeny from terminal sires have either a Merino dam (~50%) or a first cross Border Leicester x Merino dam.

From these data all animals with at least sire and dam pedigree for 2 generations and born from 2005 and later were included. Data were extracted for birth weight (Bwt), weaning weight (Wwt), post weaning weight (Pwt), post weaning fat depth (Pfat) and post weaning eye muscle depth (Pemd). All contemporary groups were transformed to a common mean within each group as is done routinely for Sheep Genetic analyses (Brown *et al.* 2007). Two analyses were undertaken using different combinations of records;

INF. Using data from all industry flocks animals with records were classified into a purebred trait if they had at least 90% breed composition of the breed of interest and animals from the Sheep CRC INF flocks into a crossbred trait if they had 50% of the breed of interest. Only the Poll Dorset breed had sufficient sires with purebred progeny in SG and first cross progeny in the INF flock to estimate the genetic correlations. The data set was reduced to all animals from the contemporary groups where the sires with progeny in both traits were represented.

SG. Using data from all flocks in Sheep Genetics, animals with records were classified into a purebred trait if they had at least 90% Poll Dorset breed composition and into a crossbred trait if they had between 25% and 75% of the Poll Dorset breed. In this dataset the breed composition of the progeny was highly variable and represented mostly crosses between terminals but also some crosses with maternal and Merino breeds.

The pedigree and breed composition was built using all ancestral information available. This resulted in pedigree files comprising between 10,835 and 132,138 animals for the INF dataset and between 379,047 and 223,424 animals for the SG dataset, depending on the trait being analysed. A summary of the number of records available for each trait in each data set is shown in Table 1.

Table 1. Summary of the number of records used for the purebred (PB) and crossbred (XB) traits, the number of sires with progeny for both traits (Sires) and the number of progeny records (Prog) from these common sires for each trait in the INF analysis (INF) and entire Sheep Genetics analysis (SG)

	Bwt		**Wwt**		**Pwt**		**Pfat**		**Pemd**	
	PB	**XB**	**PB**	**XB**	**PB**	**XB**	**PB**	**XB**	**PB**	**XB**
INF - Poll Dorset										
Rec.	117,958	3,898	120,237	3,254	107,696	3,070	91,686	2,559	90,580	2,559
Sires	118		125		126		123		123	
Prog.	27,131	3,464	28,719	3,098	25,123	2,986	21,832	2,468	21,652	2,468
SG - Poll Dorset										
Rec.	207,237	66,753	298,053	80,994	217,139	58,637	174,783	48,859	174,631	48,793
Sires	321		459		346		299		298	
Prog.	51,470	4,504	71,499	5,650	46,291	3,702	35,498	2,913	35,408	2,912

Models of analysis. Parameters were estimated in bivariate animal model analyses for each trait in ASReml (Gilmour *et al.* 2006) with purebred and crossbred performance considered as two different traits with a genetic correlation r_{PC}. For weight traits the fixed effects of contemporary group, birth type, rearing type, age of dam, and animal's age at measurement were fitted. For the carcase traits the fixed effects of contemporary group and the animal's liveweight at measurement

(as quadratic) were fitted. Contemporary group was defined as flock, year of birth, sex, date of measurement, management group subclass. A random term for the direct genetic effects was modelled for all traits. An additional random term for sire by flock year interactions was fitted for all traits and maternal effects included for the weight traits.

RESULTS AND DISCUSSION

In the INF analysis the phenotypic variances and heritabilities were both significantly higher for the crossbred traits (Table 2). This might be due to the more diverse sampling of sires and also a greater genetic diversity in the dams, of which many were lacking complete pedigree with which to account for these effects. In the analyses of the entire SG datasets the phenotypic variance and heritabilities were not significantly different between the purebred performance and crossbred traits and also agreed with previously published estimates from these data (Brown *et al.* 2015).

Table 2. Phenotypic variance (σ^2_p), direct (h^2) heritability purebred (PB) and crossbred (XB) performance for each trait and breed in the INF analysis (INF) and entire Sheep Genetics analysis (SG) (s.e. in parentheses)

		Bwt		Wwt		Pwt		Pfat		Pemd	
		PB	XB	PB	XB	PB	XB	PB	XB	PB	XB
INF	σ^2_p	0.66 (0.00)	0.76 (0.00)	26.00 (0.14)	34.31 (1.16)	34.05 (0.19)	48.27 (1.71)	NC	NC	4.53 (0.03)	5.47 (0.16)
	h^2	0.15 (0.01)	0.07 (0.03)	0.29 (0.01)	0.82 (0.10)	0.27 (0.01)	0.64 (0.09)	NC	NC	0.27 (0.01)	0.28 (0.04)
SG	σ^2_p	0.64 (0.00)	0.60 (0.00)	25.34 (0.08)	22.59 (0.14)	32.93 (0.12)	32.08 (0.24)	0.36 (0.00)	0.40 (0.00)	4.48 (0.02)	4.41 (0.04)
	h^2	0.16 (0.00)	0.20 (0.01)	0.08 (0.00)	0.10 (0.01)	0.13 (0.00)	0.14 (0.01)	0.22 (0.01)	0.24 (0.01)	0.25 (0.01)	0.28 (0.01)

NC: Analysis did not converge

Estimates of genetic correlations between purebred and crossbred performance are shown in Table 3. Taking into account the standard errors of each estimate, all correlations were not significantly different to one. This indicates that genetically the performance of animals in purebred flocks is the same as that in crossbred flocks. This also suggests that crossbred data is just as valuable as purebred data for both estimation of breeding values and development of genomic predictions for Poll Dorsets. We expected to observe lower correlations in the INF dataset compared to the SG dataset as the INF dataset was dominated by terminal x Merino crosses which are genetically more divergent crosses than terminal x terminal crosses which dominate the SG dataset, but this was not supported by our estimates.

Table 3. Genetic correlation between purebred and crossbred performance for each trait in the INF analysis (INF) and entire Sheep Genetics analysis (SG) (s.e. in parentheses)

Dataset	Bwt	Wwt	Pwt	Pfat	Pemd
INF	0.97 (0.15)	0.94 (0.16)	1.00 (0.14)	0.95 (0.10)	0.99 (0.07)
SG	0.97 (0.06)	0.99 (0.06)	0.97 (0.07)	0.92 (0.07)	0.89 (0.06)

Pfec was also analysed however there were only sufficient data available for the Poll Dorset breed when using the entire SG database, There were 46 sires with both purebred and cross bred progeny and the estimated genetic correlation was 0.92 (0.40).

Our results agree with those of Nakavisut *et al*. (2005) who estimated high correlations between purebred and crossbred performance for growth and conformation traits in pigs. Wei and van der Werf (1995) estimated genetic correlations between purebred and crossbred performance for poultry egg production traits ranging between 0.56 and 0.99. Nakavisut *et al.* (2005) observed that for reproduction traits r_{PC} was lower and in some cases significantly less than one. Thus further investigation of these correlations for wool and reproduction traits is warranted in sheep, as it would be expected that such traits have more impact from dominance effects and therefore lower purebred crossbred correlations.

CONCLUSIONS

The genetic correlations estimated between purebred and crossbred progeny performance were all very high and not significantly different to one. These preliminary estimates from industry data are consistent with those in the literature and reconfirm the use of the LAMBPLAN across breed ASBVs produced from animals with variable breed composition and also the use of crossbred animals in the genomic reference populations.

ACKNOWLEDGMENTS

This research is funded by Meat and Livestock Australia and Australian Wool Innovation, and Sheep Genetics is made possible through the support of the Australian sheep industry. The authors acknowledge the contributions of the Sheep CRC Information Nucleus flocks.

REFERENCES

Banks R.B., Brown D.J. and Field S. (2009) *Proc. Ass. Adv. Anim. Breed. Genet.* **18:** 480.

Brown D.J., Huisman A.E., Swan A.A., Graser H-U., Woolaston R.R., Ball A.J., Atkins K.D. and Banks, R.B. (2007) *Proc. Ass. Adv. Anim. Breed. Genet.* **17**: 187.

Brown D.J., Swan A.A., Gill J.S., Ball A.J. and Banks, R.B. (2015) *Anim. Prod. Sci.* (In Press)

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Ingham V.M., Fogarty N.M., Gilmour A.R., Brown D.J., Cummins L.J., Gaunt G.M., Stafford J. and Hocking Edwards, J.E. (2005) *Proc. Ass. Adv. Anim. Breed. Genet.* **16**: 227.

Nakavisut S., Crump R., Suarez M., and Graser H-U. (2005) *Proc. Ass. Adv. Anim. Breed. Genet.* **16**: 99.

Wei M. and van der Werf J.H.J. (1995) *J. Anim. Sci* **73**: 2220.

Wei M., van der Werf J.H.J. and Brascamp E.W. (1991). *J. Anim. Breed. Genet.* **108:** 262.

van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Anim. Prod. Sci.* **50:**998.

Proc. Assoc. Advmt. Breed. Genet. 21: 181-184

USING HUMP HEIGHT TO ESTIMATE BOS INDICUS CONTENT IN THE MSA BEEF GRADING MODEL WHICH PREDICTS EATING QUALITY

M. Kelly[1], J.M. Thompson[1], L.R. Porto-Neto[2], R.E. Lyons[3] and S. S. Moore[1]

[1] The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, Brisbane, QLD 4062, Australia.
[2] CSIRO Agriculture Flagship, Brisbane, QLD 4067, Australia.
[3] The University of Queensland, School of Veterinary Sciences, Gatton, QLD 4343, Australia.

SUMMARY

Bos indicus content is a key predictor of beef eating quality under Meat Standards Australia (MSA) quality grading scheme. Initially a phenotypic estimate of the proportion of *B. indicus* was used in the MSA model although more recently this has been estimated from hump height and carcass weight. The Admixture software was used to develop an estimate of *B. indicus* content using genomic information. It was demonstrated that *B. indicus* content could be accurately estimated from SNP genotype data (BB_genotype). This knowledge was used to examine the accuracy of estimating *B. indicus* content from hump height (BB_hump). The estimation of *B. indicus* content using BB_hump was found to provide a moderate accuracy of estimating *B. indicus* percentage. However, this difference in accuracy did not translate into substantial differences in the prediction of eating quality under the MSA model.

INTRODUCTION

The Meat Standards Australia (MSA) beef grading model uses commercial inputs at grading to predict beef eating quality. The MSA prediction of eating quality is based on a series of equations for individual cuts for up to five different cooking methods. The MSA model inputs include the following traits, estimated *Bos indicus* content (estBI%), whether the animal was treated with hormonal growth promotants (HGP), sex (female or steer), carcass characteristics (carcass suspension method, carcass weight, marbling and ossification scores, ribfat, and ultimate pH), and value adding effects (cooking method and days aged). One of the key animal predictors in the MSA model is estBI% (Thompson 2002, Watson *et al.* 2008). The impact of estBI% on eating quality was estimated by Watson *et al.* (2008) to be between three and 13 MQ4 score units for the different cuts in the carcase. For the MSA model development, estBI% was derived largely from pedigreed animals, or from herds of known *B. indicus* content.

When the MSA model was initially implemented nationally in 2000, estBI% was determined from the national vendor declaration (NVD) in conjunction with a physical inspection of the cattle by a trained MSA grader. Mixed lots of cattle were graded to the highest estBI% in the lot. This often necessitated redrafting mixed lots into like groups at the abattoir, which generally increased stress on the animals prior to slaughter. Hump height adjusted for carcass weight was proposed as an alternative method of assessing estBI% (Sherbeck *et al.* 1996) and this was included in the MSA model. Over time, the usage of hump height adjusted for carcass weight has increased until it is now preferred to NVDs for assessing estBI% at grading (MSA, *unpublished data*).

Given that the regression coefficients for the effect of estBI% on eating quality were largely generated from animals of known genotype it was timely to confirm the accuracy of using hump height adjusted for carcass weight compared with using animal phenotype to predict eating quality. A series of experiments performed by CRC and MSA provided the opportunity to explore relationships between estBI% estimated from hump and carcass weight (BB_Hump), genomics

(BB_Genotype) and BI% from known phenotype. It was also possible to compare the relative accuracy of BB_Hump and BB_Genotype as predictors of eating quality in the MSA model.

MATERIALS AND METHODS

A series of datasets were used to investigate different objectives. The first objective was to develop a estimate of *B. indicus* content from genomic SNP panels and then test this estimate using independent data. The second objective was to evaluate the relationship between estimated B. indicus content (estBI%) predicted from hump height and carcass weight (BB_Hump) and estBI% from genomic information (BB_Genotype). Finally, the accuracy of using either BB_Hump, or BB_Genotype along with other MSA input traits to predict eating quality, was assessed.

Development of BB_Genotype estimate. To develop an estimate of *B. indicus* content (BB_genotype) and evaluate its efficacy for prediction of BI content, the CRC III genotype and phenotype databases were used. To build the estimate, a training set of 5,650 animals and a validation set of 9,734 animals were selected from the total data set. Within breeds, animals were randomly assigned to training and validation groups. The diversity of breeds in the CRC III database meant there was a wide range of breeds and crossbred animals used to test the accuracy and precision of the *Bos indicus* content estimates. A subset of 5,817 markers that were common across all Illumina 10k, HD50k and 700k genotyping platforms were selected in the prediction equation for BB_Genotype.

Admixture software was used to develop estimates of BB_Genotype from SNP data (Alexander *et al.* 2009, Alexander and Lange 2011). The animals selected as training animals were coded as either *Bos taurus* (BT) or *Bos indicus* (BI) and the supervised option was used. The animals set coded as BT included Angus (n=2,000), Murray Grey (n=200), Shorthorn (n=500), Hereford (n=500), Limousin (n=50) and Charolais (n=400). There were 2,000 Brahmans used in the analysis as the BI reference. This program has been used previously (Porto Neto *et al.* 2014) to estimate breed composition in beef cattle.

The relationship between BB_Hump and BB_Genotype. This relationship was assessed using three data sets. Firstly, the CRC II data which comprised 1,012 animals that had been slaughtered and MSA graded. Secondly, the long distance transport (LDT) data set (Polkinghorne *et al.* 2013) which comprised 343 cattle, and lastly, 50 animals from a Rigor Temperature Experiment (RTE) (J Thompson, unpublished data). A simple linear regression was used to estimate the relationship between BB_Hump and BB_Genotype.

Prediction of eating quality using either BB_Hump or BB_Genotype in the MSA model. The LDT and RTE data sets had consumer eating quality on striploin samples. For both data sets regression models to predict eating quality (MQ4 score), included MSA input traits (carcass weight, marbling and ossification scores, ribfat and ultimate pH) along with terms for either BB_Hump or BB_Genotype. A multiple regression was used to assess the relationship between BB_Hump and BB_Genotype following adjustment for components routinely considered in the MSA eating quality prediction model.

RESULTS AND DISCUSSION

The genomic estimate of Brahman content using SNP data was shown to be closely related to Brahman content from pedigree (R^2=98%). This was slightly higher than the estimate of Frkonja *et al.* (2012) who was able to explain approximately 94% of the breed composition. However, in the earlier study, the breeds comprised Simmental and Red Holstein Friesian which were much less divergent than in the current data set. In addition, the study by Frkonja *et al*. used a much smaller training data set (495 cattle). Likewise, Kuehn *et al.* (2011) was able to explain between 77% and 92% of the variation in breed composition within *Bos taurus* beef breeds.

Table 1. The coefficient of determination (R^2), residual standard deviation (RSD), intercept and slope for the relationship between BB_Genotype and BB_Hump for three data sets.

Data set	Range BB_Genotype	R^2	RSD	Intercept	Slope
CRC II	40-100%	0.44	25.3	0.08 (2.07)	0.84 (0.02)
RTE	0-100	0.70	17.8	22.2 (3.06)	1.43 (0.14)
LDT	0-100	0.40	15.5	44.8 (1.21)	0.56 (0.04)

To examine the relationship between hump height and *B. indicus* content three data sets were used (Table 1). Within all three data sets there was a positive relationship between BB_Genotype and BB_Hump, the coefficient of determination ranging from 40 to 70%. The residual standard deviation indicated that the error in predicting BB_Genotype from BB_Hump was similar for the two RTE and LDT datasets, but larger for the CRC II dataset. The slopes of the different regressions indicated that BB_Genotype was under or overestimated in the different data sets. Thus, there may be scope to adjust the equation used to predict *BB%* from BB_Hump in order to reduce bias.

Table 2. F ratios for input traits used to predict palatability (MQ4) using data from the rigor temperature (RTE) and the long distance transport (LDT) experiments.

MSA traits	F Ratio			
	RTE		LDT	
	Model 1	Model 2	Model 1	Model 2
BB_Genotype	18.52	-	15.33	-
BB_Hump	-	12.62	-	10.04
Steak Position	7.19	7.09	13.17	16.05
Sex	1.49	1.89	-	-
Hang	26.32	26.13	-	-
Hang*position	1.74	1.72	-	-
HSCW	0.51	0.65	2.68	2.4
Ribfat	0.73	0.57	2.28	1.76
Ossification	2.11	2.11	1.92	2.49
Marbling	26.64	24.91	23.62	27.04
Ultimate pH	3.22	2.8	1.08	0.88
Days aged	15.57	16.48	-	-
HGP	-	-	0.18	0.6
Residual standard deviation	10.46	10.55	10.6	10.65
R^2	39.38	38.16	24.50	23.22

Model 1 includes BB_Genotype and Model 2 includes BB_Hump. Degrees of freedom (DF) for Rigor temperature 1,281 for all terms except position (2,281), hang (2,281), position* hang (4281) and days aged (3281)

DF for LDT was 1312 for all terms except position (2,212)

When adjusted for other terms in the MSA model the regression coefficients for the two estimates of *B. indicus* content differed slightly. When *B. indicus* content was estimated by BB_Hump the regression coefficient indicated that an increase in *B. indicus* content from 0 to 100% resulted in a decrease of 14 MQ4 units in palatability. By contrast when BB_Genotype was used the decrease was only 9 MQ4 units.

Using two MSA datasets the accuracy of predicting eating quality (MQ4) was similar regardless of whether it was estimated using BB_Genotype, or BB_Hump (Table 2). The use in isolation of BB_Genotype or BB_Hump had a large effect on the total F ratio (data not shown). However, there was little difference in the overall percentage of variation explained when using either BB_Hump or BB_Genotype to adjust for BB% under the MSA model. The coefficient of determination dropped by approximately 1% in both cases. In both experiments, the F ratios were slightly higher when estBI% was predicted from BB_Genotype compared with using BB_Hump. As the variation explained using all the MSA inputs did not change substantially, the variation that should have been due to BB% under the BB_Hump term was partitioned across other terms in the model. In the case of LDT this variation was picked up by marbling score possibly due to the correlation between marbling score and BB% in this data set. In RTE it was less clear which individual terms accounted for the difference in using BB_Genotype or BB_Hump.

IMPLICATIONS

Using data from a number of MSA experiments, BB_Hump predicted BB_Genotype with reasonable accuracy, although in the different data sets there was a tendency to either over or underestimate BB_Genotype. This could be corrected by adjusting the formula used to convert hump height to BB_hump, or by using a genomic estimate of *B. indicus* content. When used in a regression model with other MSA inputs, both BB_Hump and BB_Genotype were similar in their ability to predict consumer eating quality.

ACKNOWLEDGEMENTS

Funding for this project was provided by Meat and Livestock Australia (MLA) project BNP.1420.

REFERENCES

Alexander, D. H. and Lange, K. (2011) *BMC Bioinformatics* **12**: 246.

Alexander, D. H., Novembre, J. and Lange, K. (2009) *Genome Res* **19**: 1655.

Frkonja, A., Gredler, B., Schnyder, U., Curik, I. and Solkner, J. (2012) *Anim. genet.* **43**: 696.

Kuehn, L. A., Keele, J. W., Bennett, G. L., McDaneld, T. G., Smith, T. P., Snelling, W. M.,Sonstegard, T. S. and Thallman, R. M. (2011) *J Anim Sci* **89**: 1742.

Porto-Neto L.R., Reverter A., Prayaga K.C., Chan E.K., Johnston D.J., Hawken R.J., Fordyce G., Garcia J.F., Sonstegard T.S., Bolormaa S., Goddard M.E., Burrow H.M., Henshall J.M., Lehnert S.A. & Barendse W. (2014) *PLoS ONE* **9**: e113284.

Thompson, J. M. (2002) *Meat Science* **62**: 295.

Watson, R., Polkinghorne, R. and Thompson, J. (2008) *Aust. J. of Exp. Agric.* 48: 1368.

Polkinghorne R, Thompson JM, Philpott J (2013). Final Report Project B.NBP.0722, Meat and Livestock Australia, Sydney.

Proc. Assoc. Advmt. Breed. Genet. 21: 185-188

ARE HIGH AUSTRALIAN PROFIT RANKING SIRES BEST IN ALL HERDS? FINDINGS FROM THE FEEDING THE GENES PROJECT

J.M. Morton[1], R.R. Woolaston[2], P. Brightling[3], S. Little[4], K.L. Macmillan[5], J.E. Pryce[6] and G.J. Nieuwhof[6,7]

[1]Jemora P/L, PO Box 2277, Geelong, VIC 3220
[2]30 Airlie Rd, Pullenvale, QLD 4069
[3]Harris Park Group, Level 2, Swann House, 22 William Street, Melbourne, VIC 3000
[4]Capacity[+] Ag Consulting, 350 Mitchell Road, Kialla, VIC 3631
[5]Department of Veterinary Science and Agricultural Science, The University of Melbourne, Princes Highway, Werribee, VIC 3030
[6]Biosciences Research Division, Department of Economic Development, Jobs, Transport & Resources, Agribio, Bundoora, VIC 3083
[7]Australian Dairy Herd Improvement Scheme, Level 2, Swann House, 22 William Street, Melbourne, VIC 3000

SUMMARY

The effects of increasing Australian Profit Ranking (APR) were assessed in 5 herd feeding systems and at various levels of milk production. In total, 505 herds and 250,857 and 43,941 lactations for Holstein and Jersey cows, respectively, were used for analyses. Effects of sire APR on milk yield variables were positive in all feeding systems and at all herd average solids per cow levels. Effects were similar for the most commonly used feeding systems but were approximately twice as large in herds with a total mixed ration feeding system than in low bail feeding herds. Cows with higher sire APRs were just as likely or more likely to recalve by 20 months as cows with lower genetic merit. Thus selecting high APR sires had benefits in all feeding systems, supporting the use of the same APR across all of these feeding systems. Herd managers using artificial breeding should select high index sires with an appropriate semen price and Australian Breeding Values that are aligned with the breeding objectives for their herd.

INTRODUCTION

The APR, introduced by the Australian Dairy Herd Improvement Scheme (ADHIS) in 2001 as a national selection index for dairy cattle, was most recently revised in 2009 (Pryce *et al*. 2009), and replaced in April 2015 with a new economic breeding index, the Balanced Performance Index (BPI). At the same time, 2 additional breeding indices were also introduced: the Health Weighted Index (HWI) and the Type Weighted Index (TWI) (Byrne *et al.* 2015; Martin-Collado *et al.* 2015). These 3 new breeding indices are closely correlated with the APR (correlation coefficients 0.98, 0.96 and 0.95, respectively (Nieuwhof G, personal communication).

The Australian dairy industry is characterised by a diverse range of feeding systems and the Australian Breeding Values used to calculate APRs are based on animal performance data pooled across Australian herds across all feeding systems. However, some advisors and farmers have questioned the validity of the APR for specific situations i.e. they ask whether there is an important genotype by environment interaction (G*E). Several trials assessing genetic merit by feeding interactions have been conducted in research herds including Kennedy *et al*. (2003), Beerda *et al*. (2007) and Fulkerson *et al*. (2008), and numerous large scale cohort studies have compared cows of varying genetic merit in commercial herds with various environments. However, only a few of these latter studies compared effects of varying genetic merit between different feeding systems; recent examples include Kearney *et al*. (2004a, 2004b) and Ramirez-Valverde *et al*. (2010). No such studies have been conducted in Australia.

Milk production and cow longevity are important to herd managers. To describe 'lasting ability', recalving by 20 months can be used to collectively describe short to medium term reproductive performance, culling and death. Thus, G*E effects on recalving by 20 months are of interest as they would impact on cow survival, lifetime milk yields, herd culling policy and replacement rates.

Milk production per cow is generally lower where the feeding system consists of pasture and conserved fodder with low concentrate use, and is much higher in herds using the total mixed ration feeding system. Herd average milk yield is readily calculated with routinely collected milk recording data whereas feeding system data are not routinely collected. When studying G*E or assessing sires in different environments, it would be simpler to define environment as herd average milk yield than feeding system. Accordingly, it was also important to assess whether feeding system is a surrogate 'environment' for herd average milk yield when assessing G*E. This could also inform the nature of any interactions detected.

The major aims of the project were: a) to estimate the effects of APR on milk production, and recalving by 20 months in cows in commercial Australian dairy herds using various feeding systems; b) to ascertain whether these effects differ substantially between herds with different feeding systems; and c) to assess whether feeding system is a surrogate environment for herd average solids (i.e. fat plus protein) per cow when assessing G*E.

MATERIAL AND METHODS

In 2012, all herds in which at least 30 Holstein cows calved in 2011 and/or at least 30 Jersey cows calved in 2011 were selected from the ADHIS database. Herds with less than 50 cows calved in 2011 were excluded. Letters were sent to managers of the remaining 2016 herds asking them to complete a simple herd data questionnaire to identify their herd's feeding system. In total, 505 herds provided data suitable for analyses and cow and lactation data for these herds were obtained from ADHIS. From these herds, 250,857 and 43,941 lactations for Holstein and Jersey cows, respectively, were used for analyses. Each cow's sire's APR was as estimated by ADHIS on 20th August, 2012. Each lactation was classified as having been followed by another calving within 20 months or not.

For 2008, 2009, 2010 and 2011, each herd's feeding system was classified as follows: low bail (grazed pasture, fed other forages and fed ≤1t grain/concentrates in parlour during milking annually/cow); moderate/high bail (grazed pasture, fed other forages and fed >1t grain/concentrates per cow in parlour during milking); partial mixed ration (a portion of the ration was fed on a feed pad using a mixer wagon and cows are fed pasture for at least 9 months of the year); hybrid (a portion of the ration was fed on a feed pad using a mixer wagon and cows are fed pasture for 2-8 months of the year); and total mixed ration (cows are fed pasture for no more than 1 month of the year). These definitions were specified based on a scheme developed by Dairy Australia (Dairy Australia 2015). Herd average solids per cow were calculated for each herd-year as the averages of each cow's 305-day fat plus protein yields.

For all analyses, the unit of analysis was the individual lactation. Phenotypic relationships between sire APR and 305-day milk yield variables were assessed using multilevel linear models; relationships between sire APR and recalved by 20 months were assessed using logistic models with herd fitted as a random effect.

Additional analyses were conducted with ASReml (Gilmour *et al*. 2009) using a genetic model to estimate genetic correlations between feeding systems for 305-day milk volume, with a subset of the data containing 60,532 first-lactation Holstein records from 3136 herd-year-season (HYS) combinations across 439 herds. Of the 2293 sires, approximately 1/2 (1131) had daughters in just 1 feeding system while only 89 had daughters in all 5 feeding systems. Most sires (87%) had fewer than 20 daughters in any feeding system and only 5% had moderately-sized families (>19

daughters) in more than 1 feeding system. Thus, the power of this data set for evaluating genetic performance in more than 1 feeding system was not strong. For analysis, base ancestors were assigned to 1 of 58 genetic groups, HYS was fitted as a fixed effect, and separate residual and sire variances were fitted for each of the 5 feeding systems, with 3 alternative structures for the latter: diagonal, correlated (uniform), and factor analytic. A second genetic model was tested using random regression, with the average milk volume of each HYS as a simple (linear) environmental descriptor instead of feeding system.

RESULTS AND DISCUSSION

Effects on milk production. For Holstein cows, effects of sire APR on milk production variables were positive in all feeding systems but differed by feeding system (Table 1). They were approximately twice as large in total mixed ration feeding system herds compared with low bail feeding herds. However, effects were more similar for the more commonly used feeding systems (low bail, moderate to high bail, and partial mixed ration feeding systems). Effects of sire APR on milk volume and protein yield also differed by herd average solids per cow. Effects were positive at all herd average solids per cow levels. However, no such interaction was evident for fat yield.

Table 1. Estimated effects*of cow's sire's APR on 305-day milk production for lactations from Holstein cows by feeding system (95% CI)

Milk production variable	Feeding system				
	Low bail	Moderate to high bail	Partial mixed ration	Hybrid	Total mixed ration
Milk volume (l)	56.2 (40.9 to 71.5)	68.0 (60.4 to 75.6)	53.7 (39.8 to 67.7)	79.7 (58.8 to 100.6)	109.9 (75.1 to 144.8)
Fat yield (kg)	2.6 (2.0 to 3.2)	2.5 (2.2 to 2.8)	1.5 (1.0 to 2.0)	3.5 (2.7 to 4.3)	5.7 (4.4 to 7.1)
Protein yield (kg)	2.6 (2.1 to 3.1)	3.4 (3.2 to 3.6)	2.9 (2.5 to 3.4)	4.0 (3.3 to 4.6)	5.1 (4.0 to 6.2)

*Coefficients represent estimated change in milk production variable per 50 unit increase in the cow's sire's APR; coefficients were adjusted for maternal grandsire's APR and age at calving; herd and cow within herd were fitted as random effects

For milk volume and protein yield, the interaction between APR and feeding system was largely accounted for by interaction between APR and herd average solids per cow. In contrast, the interaction between APR and feeding system for fat yield was not accounted for by interaction between APR and herd average solids per cow. These results indicate that the biological determinants of G*E for fat yield differ from those for milk volume and protein yield. Features of feeding systems determine APR effects on fat yield. In contrast, factors associated with herd average milk yield determine G*E effects of APR on milk volume and protein yield.

For Jersey cows in herds using low and moderate to high bail feeding systems and partial mixed ration feeding, increases in sire APR increased milk volume, and fat and protein yields. Increases in milk volume, and fat and protein yield were smaller for the low bail feeding system than for the other 2 feeding systems.

In the genetic analyses of milk volume, a uniform structure was found to be statistically the best fit for the genetic correlation between feeding systems, with an estimate of 0.81±0.06. Data structure limited the power to test more complex correlation structures. The random regression model revealed a correlation of 0.81±0.08 between the slope of the regression and its intercept (i.e. between responsiveness to production level and genetic merit). Collectively these indicate that

genetic expression for milk volume was strongly correlated across all 5 feeding systems, and that the superiority of bulls with high ABV tended to increase as the herd's average milk volume increased.

Effects on recalving by 20 months. Cows with higher sire APRs were just as likely (if not more likely) to recalve by 20 months as cows with lower genetic merit. Estimated effects of increasing APR on whether a cow recalved by 20 months were weakly positive across all except the total mixed ration feeding system, and across all herd milk yield categories; effects were stronger in herds with higher herd average solids per cow.

CONCLUSIONS

In all feeding systems, the daughters of higher APR sires produced more milk and were just as likely (if not more likely) to last in the herd as daughters of lower APR sires. This shows that herd managers do not need to feed high rates of supplements to benefit from selecting high APR sires and that the daughters of high APR sires are likely to last as long or longer in the herd than daughters of lower APR sires.

The magnitude of benefits of greater genetic merit varies between feeding systems (i.e. there was an interaction between genetic merit and feeding system). The response from selecting high APR sires was realised in all systems but was greater in herds using more intensive feeding systems (hybrid and total mixed ration). The biological determinants of G*E for fat yield differ from those for milk volume and protein yield. Features of feeding systems determine APR effects on fat yield. In contrast, determinants associated with herd average milk yield determine APR effects on milk volume and protein yield.

Given the very close correlations between APR and each of the 3 new indexes, similar conclusions should apply for these. In summary, herd managers using artificial breeding should select high BPI, HWI or TWI sires with an appropriate semen price and Australian Breeding Values that are aligned with the breeding objectives for their herd.

REFERENCES

Beerda B., Ouweltjes W., Sebek L.B.J., Windig J.J. and Veerkamp R.F. (2007) *J. Dairy Sci.* **90**: 219.

Byrne T., Martin-Collado D., Santos B., Amer P., Pryce J. *et al*. (2015) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21**: (in press).

Dairy Australia (2015) Australia's 5 main feeding systems. [Online]. Available at http://www.dairyaustralia.com.au/Pastures-and-Feeding/Feeding-systems/Flexible-feeding-systems.aspx (accessed 16 April 2015).

Fulkerson W.J., Davison T.M., Garcia S.C., Hough G., Goddard M.E. *et al*. (2008) *J. Dairy Sci.* **91**: 826.

Gilmour A.R., Gogel B.J., Cullis B.R., and Thompson R. (2009) ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK (www.vsni.co.uk)

Kearney J.F., Schutz M.M., Boettcher P.J. and Weigel K.A. (2004a) *J. Dairy Sci.* **87**: 501.

Kearney J.F., Schutz M.M. and Boettcher P.J. (2004b). *J. Dairy Sci.* **87**: 510.

Kennedy J., Dillon P., Delaby L., Faverdin P., Stakelum G. et al. (2003) *J. Dairy Sci.* **86**: 610.

Martin-Collado D., Byrne T. J., Amer P. R., Santos B. F. S., Axford M. *et al.* (2015) *J. Dairy Sci.* **98**: 4148.

Pryce J.E., Van der Werf J.H.J., Haile-Mariam M., Malcolm B. and Goddard M.E. (2009) *Proc. Assoc. Advmt. Anim. Breed. Genet.***18**: 143.

Ramirez-Valverde R., Peralta-Aban J.A., Nunez-Dominguez R., Ruíz-Flores A., García-Muñiz J.G. *et al*. (2010) *Animal* **4**: 1971.

Proc. Assoc. Advmt. Breed. Genet. 21: 189-192

SELECTION FOR REDUCED MATURE COW WEIGHT WILL REDUCE BODY CONDITION UNLESS ACCOMPANIED BY SELECTION FOR INCREASED FAT

W.S. Pitchford[1], S.J. Lee[1], K.A. Donoghue[2] and B.J. Walmsley[3]

[1]School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy SA 5371
[2] Department of Primary Industries, Trangie Agricultural Research Centre, Trangie NSW 2823
[3] Animal Genetics and Breeding Unit*, University of New England, Armidale NSW 2351

SUMMARY

A large number of Angus cows (3,768) were ultrasound scanned for eye muscle area, rib and rump fat depth and intramuscular fat both as yearlings and at weaning of their first calf. They also had weight, height and body condition recorded. Response to selection for a number of scenarios was evaluated following current industry index weightings. Single trait selection for reduced cow weight at time of weaning her calf would result in shorter cows with less fat, muscle and condition. However, selection indices that include some positive weighting on carcass fat (low weighting on subcutaneous fat depth or large weighting on intramuscular fat) would result in increased body condition of cows even when adjusted for changes in weight or mature size. Recently updated Angus selection indices do have positive weightings on both subcutaneous and intramuscular fat.

INTRODUCTION

The Cooperative Research Centre for Beef Genetic Technologies (Beef CRC) included a large Maternal Productivity project. The Project was motivated by seedstock breeder concerns that the body composition of cows is changing in response to selection for feedlot performance which is to the detriment of the breeder herd, especially during seasons with reduced feed availability (Pitchford *et al.* 2015). These concerns were captured in a social science study of seedstock breeders (Lee *et al.* 2015a). Implicit in this concern is a lack of confidence in selection indexes at the time. The specific concern addressed in this paper is that selection for low mature weight which is designed to account for feed costs of cows will result in cows that are of lower body condition rather than being of "more moderate frame".

MATERIALS AND METHODS

One part of the Maternal Productivity Project was scanning cows that were already recorded for existing BREEDPLAN traits. The results reported herein are for 3,768 Angus cows that were measured at the time of weaning of their first calf as reported by Donoghue *et al.* (2015). The traits measured on the cows were weight (WT, kg), hip height (HT, cm), condition score (CS; 1-5 scale; Graham 2006), ultrasound scanned eye muscle area (EMA, cm^2), fat depth at the 12/13th rib and P8 rump sites (RIB and P8, mm) and intramuscular fat (IMF, %). These cows (mature, M) were also measured as yearling (Y) heifers for 400 day WT, EMA, RIB, P8 and IMF for routine estimation of BREEDPLAN EBVs.

The effects on cow weight and composition were assessed using selection index theory (Hazel 1943). This was based on correlations rather than covariances as the relative changes were considered more important than the absolute changes. The relative value of 12 traits was a vector of weights, **v**. The vector of 7 selection weights (**b**) were calculated as:

$$\mathbf{b} = \mathbf{G_{11}}^{-1}\mathbf{G_{12}}\mathbf{v}$$

* AGBU is a joint venture of NSW Dept. of Primary Industry and the University of New England

where G_{12} is a 7x12 matrix of correlations between the 7 cow traits and the 12 potential traits (7 cow plus 5 yearling) and G_{11}^{-1} is the inverse of the correlation matrix between the 7 cow traits.

The variance of the index (σ_I^2) and the vector of response to selection in the 7 cow traits (**R**) were calculated as:

$$\sigma_I^2 = \mathbf{b'G_{11}b}$$

$$\mathbf{R} = \mathbf{b'G_{11}}/\sigma_I$$

Note that responses are in standard deviation units and should only be considered relative as they are scaled by the standard deviation of the index.

The genetic correlations between cow traits and yearling heifer traits are taken from Donoghue *et al.* (2015, Tables 4 and 8) and are presented in Table 1. The cow traits used were from time of weaning of their first calf because this time point had the most data and were very highly genetically correlated (generally >0.9) with traits recorded after this time point. Thus, the measurements used herein are assumed to represent cow condition at later ages. The correlations between heifer traits and cow traits were only reported for the same trait across time, so the correlations between heifer (yearling, Y) trait x and cow (mature, M) trait y were estimated from the same data set but have not been published previously.

Table 1. Heritabilities, genetic standard deviations and genetic correlations between 5 heifer (Y) and 7 cow (M) traits (G_{12}')

Trait	h^2	σ_A	MWT	MP8	MRIB	MEMA	MIMF	MCS	MHT [A]
MWT	0.45	22.5	1	0.22	0.19	0.53	0.18	0.39	0.70
MP8	0.44	1.33	0.22	1	0.96	0.46	0.71	0.87	-0.15
MRIB	0.46	0.97	0.19	0.96	1	0.45	0.73	0.87	-0.14
MEMA	0.26	3.07	0.53	0.46	0.45	1	0.33	0.65	0.17
MIMF	0.32	0.84	0.18	0.71	0.73	0.33	1	0.71	0.07
MCS	0.14	0.17	0.39	0.87	0.87	0.65	0.71	1	-0.25
MHT [A]	0.57	2.61	0.70	-0.15	-0.14	0.17	0.07	-0.25	1
YWT	0.31	16.6	0.71	-0.07	-0.04	0.03	-0.05	-0.01	0.57
YP8	0.46	1.37	-0.21	0.49	0.47	-0.10	0.28	0.39	-0.24
YRIB	0.45	0.98	-0.16	0.45	0.57	0.05	0.37	0.43	-0.21
YEMA	0.35	3.49	0.15	-0.04	-0.04	0.59	0.04	0.11	0.15
YIMF	0.29	0.81	-0.26	0.23	0.32	-0.04	0.65	0.16	-0.12

[A] Actually measured prior to calving but extremely repeatable so assumed same trait.

Thirteen scenarios were tested for the effect of selection pressure on cow weight and body composition. The relative weightings used in the scenarios were based on a subset of those in the current Angus Breeding Index (Angus Australia 2014). The traits of importance herein are yearling weight, P8 fat, eye muscle area, intramuscular fat and mature cow weight and the relative emphasis has been assumed to be +19%, +6%, +2%, +11% and -4% respectively. In the Angus Breeding Index it is actually 600d weight that is +19% and 400d weight is only +3%. However, for the purposes of modelling herein, it was assumed that yearling weight was the trait with the greatest selection pressure at +19%. All current Angus indexes are highly correlated with each other so the choice of which specific index to use is unlikely to impact on the conclusions herein.

The first series of scenarios (Table 2) were based on single trait selection for decreasing mature weight, increasing cow condition score or decreasing cow height (frame). Selection using combinations of these cow traits were then tested in scenarios 4 and 5. Scenario 6 assumed the sole focus was to “bend the growth curve” with high yearling and low mature weight. Various carcass quality measures were added to this in the remaining scenarios with scenario 13 being interpreted as similar to the Angus Breeding Index.

Table 2. Weighting on objective traits for selection scenarios tested.

Scenario		YWT	YP8	YEMA	YIMF	MWT	MCS	MHT
1.	↓MWT					-4		
2.	↑MCS						+4	
3.	↓MHT							-4
4.	↓MWT+↑MCS					-4	+4	
5.	↓MHT+↑MCS						+4	-4
6.	↑YWT+↓MWT	+19				-4		
7.	↑YWT+↓MWT+↑YP8	+19	+6			-4		
8.	↑YWT+↓MWT+↑YEMA	+19		+2		-4		
9.	↑YWT+↓MWT+↑YIMF	+19			+11	-4		
10.	↑YWT+↓MWT+↑YEMA+↑YP8	+19	+6	+2		-4		
11.	↑YWT+↓MWT+↑YEMA+↓YP8	+19	-6	+2		-4		
12.	↑YWT+↓MWT+↑YEMA+↑YIMF	+19		+2	+11	-4		
13.	↑YWT+↓MWT+↑YEMA+↑YP8+↑YIMF	+19	+6	+2	+11	-4		

RESULTS AND DISCUSSION

In this project, mature cow condition score was highly genetically correlated with mature fat depth (0.87) and strongly correlated with MEMA and MIMF (0.65, 0.71; Table 1). Some producers had concerns about cows that could be the same weight, but some are tall with low condition and others are of modest stature with high condition score. Height and MEMA were more strongly correlated with MWT (0.70 and 0.53, respectively) than fat (0.18-0.22) and condition score (0.39) with MWT. Mature cow condition score was lowly correlated with yearling measurements of WT, EMA and IMF and moderately correlated with fat depth (P8 and RIB).

Selection for solely decreased mature weight resulted in cows that had less fat, less muscle and were shorter (Scenario 1, Table 3). This result supports the concerns of some breeders that selection pressure for lower mature weight is associated with cows that have poorer condition. This is important as cow energy reserves influence reproductive performance (Osoro and Wright 1992) and, therefore time retained in the herd. A potential strategy could be to select for increased cow condition rather than weight *per se*. This resulted in cows that were heavier but shorter and, as expected, had more muscle and fat (Scenario 2, Table 3). Height was negatively correlated with condition (Table 1) so selecting for shorter cows resulted in decreased weight (assumed favourable for reducing feed requirements) and increased condition (also considered favourable; Scenario 3, Table 3). Viewed simply, selection for reduced height could be more favourable than selecting for decreased weight. Scenarios 4 and 5 demonstrate that placing a positive weight on MCS ensures greater increase in cow condition. However, selection for cow traits needs to be in the context of a broader selection index.

Selection with a strong emphasis on increased yearling weight but with lowered emphasis on reducing mature cow weight (Scenario 6) resulted in the opposite response to Scenario 1 where cow weight and height increased but all measures of condition still decreased. Cow body condition is a function of both muscle and fat (Graham 2006). However, as heifer EMA and IMF were lowly correlated with cow condition (Table 1), only scenarios with a positive emphasis on fat

depth resulted in increased cow condition. Scenarios 9 and 12 with significant emphasis on increasing IMF did result in maintained MCS..

Table 3. Correlated changes in cow weight and composition resulting from various selection scenarios and expressed as multiples of σ_A.

	Scenario	MWT	MP8	MRIB	MEMA	MIMF	MCS	MHT
1.	↓MWT	-1.00	-0.22	-0.19	-0.53	-0.18	-0.39	-0.70
2.	↑MCS	0.39	0.87	0.87	0.65	0.71	1.00	-0.25
3.	↓MHT	-0.70	0.15	0.14	-0.17	-0.07	0.25	-1.00
4.	↓MWT+↑MCS	-0.55	0.59	0.62	0.11	0.48	0.55	-0.86
5.	↓MHT+↑MCS	-0.20	0.65	0.64	0.30	0.40	0.79	-0.79
6.	↑YWT+↓MWT	0.79	-0.18	-0.13	-0.13	-0.14	-0.15	0.67
7.	↑YWT+↓MWT+↑YP8	0.68	0.06	0.11	-0.18	0.00	0.05	0.55
8.	↑YWT+↓MWT+↑YEMA	0.84	-0.20	-0.14	-0.03	-0.14	-0.13	0.71
9.	↑YWT+↓MWT+↑YIMF	0.63	0.03	0.19	-0.19	0.52	0.00	0.64
10.	↑YWT+↓MWT+↑YEMA+↑YP8	0.75	0.06	0.11	-0.09	0.01	0.07	0.61
11.	↑YWT+↓MWT+↑YEMA+↓YP8	0.88	-0.42	-0.35	0.02	-0.26	-0.31	0.78
12.	↑YWT+↓MWT+↑YEMA+↑YIMF	0.69	0.02	0.19	-0.08	0.55	0.02	0.69
13.	↑YWT+↓MWT+↑YEMA+↑YP8+↑YIMF	0.49	0.27	0.41	-0.12	0.63	0.22	0.48
14.	13 Conditional on MWT	0.00	0.25	0.39	-0.26	0.61	0.15	0.24

Lee *et al.* (2015b) distinguished between fat depth and fatness, which they defined as fat depth with weight fitted as a covariate. Scenario 13 resulted in increases in all cow traits except EMA with the relative increase in height being smaller than weight and carcass composition. This raises the question as to whether the carcass traits just increased with weight or whether cow composition was genuinely improved. In theory, condition score should address this, but given it is subjective it seemed sensible to assess changes independent of those in weight (Scenario 14). The result was a greater reduction in EMA (-0.26), reduced change in height (0.24) and similar changes in fat traits. In conclusion, as yearling IMF and EMA were lowly correlated with cow condition score, a small positive weighting on yearling fat depth or large weighting on IMF is required to avoid selection leading to reduced body condition of cows below that necessary for production.

ACKNOWLEDGEMENT

This research was primarily funded by Meat and Livestock Australia through the Cooperative Research Centre for Beef Genetic Technologies.

REFERENCES

Angus Australia (2014) http://www.angusaustralia.com.au/images/Detailed_Guide_to_Angus_Selection_Indexes.pdf Accessed April 2015.
Donoghue, KA, Lee, SJ, Parnell, PF and Pitchford WS (2015) *Anim. Prod. Sci.* Accepted.
Hazel, LN (1943) The genetic basis for constructing selection indexes. *Genetics* 28, 476-490.
Graham JF (2006) Condition scoring of beef cattle. Department of Primary Industries Victoria. Available at http://www.depi.vic.gov.au/agriculture-and-food/livestock/beef/handling-and-management/condition-scoring-of-beef-cattle Accessed April 2015
Lee SJ, Nuberg IK and Pitchford WS (2015a) *Anim. Prod. Sci.* Accepted.
Lee SJ, Donoghue KA and Pitchford WS (2015b) *Anim. Prod. Sci.* Accepted.
Osoro K and Wright IA (1992) *J. Anim. Sci.* 70,1661-1666.
Pitchford WS, Accioly JA, Banks RG, *et al.* (2015) *Anim. Prod. Sci.* Accepted.

Proc. Assoc. Advmt. Breed. Genet. 21: 193-196

ESTIMATES OF GENETIC PARAMETERS FOR SEASONAL WEIGHT CHANGES OF BEEF COWS

Karin Meyer[1] and Ian G. Colditz[2]

[1]Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351
[2]CSIRO Agriculture, Armidale, NSW 2350

SUMMARY

Genetic parameters were estimated for seasonal body weight changes of cows and weaning weight of their calves in two beef herds run at pasture in a Mediterranean climate. Heritability estimates for weight changes were low. Cows predisposed to lose more weight were also likely to gain more weight, and larger cows had greater genetic potential for weight changes. Low to moderate genetic and permanent environmental correlations indicated that cows with greater seasonal weight changes weaned heavier calves, due in part to the genetic association between weaning weight and cows' mature body weight. Results indicate that in this environment, scope to select for heavy weaning weight without penalty to cow body weight during periods of seasonal feed scarcity is limited.

INTRODUCTION

Pasture based livestock production in Australia is affected by extreme seasonal variation which results in an annual pattern of weight gains and losses, depending on feed availability. Resource allocation theory posits that an ability to wean a heavy calf with little penalty to her own body weight should provide a cow with 'robustness' to environmental challenges. In spite of increasing interest in robust cows, few studies reporting genetic parameters for body weight changes are available. We examine patterns of variation for seasonal weight changes of beef cows and the relation to growth of their calves using data from a selection experiment in Western Australia.

MATERIAL AND METHODS

Data originated from the Wokalup selection experiment which comprised two herds of approximately 300 cows each, Polled Herefords (HEF) and a synthetic breed, the so-called Wokalups (WOK); details are given by Meyer *et al.* (1993). Except during calving, all animals were weighed on a monthly basis. Production was entirely pasture based and governed by a Mediterranean climate with winter rains and summer droughts, i.e. feed growth in winter and spring and subsequent dearth in summer and autumn. Calving took place mainly in April and May and calves were weaned, depending on the season, in late November or December. This resulted in strong seasonal variation in body weight, with cows usually at their top weight in January and lowest weight in June.

Analyses. Traits considered were cow body weights in January (JAN) and June (JUN), weight changes from January to June (LOSS) and June to the following January (GAIN) and calf weaning weight (WW), disregarding observations for cows more than 8 years old and WW records for calves not raised by their genetic dam. Characteristics of the data are summarised in Table 1.

Data were analysed fitting a random regression (RR) model, treating WW as a characteristic of the cow. Fixed effects comprised a quadratic regression on age of cow (in years) and contemporary groups, defined as year-paddock classes for the cow traits and year-paddock-sex of calf classes for WW. Other effects for cow traits included month of calving, assigning a code of '0' for cows without calves recorded. For WW, birth type (single or twin; <2%) as well as a within sex linear regression on age at weaning, were fitted. Random effects were additive genetic (G) and permanent environmental (PE) effects of the cow, modelling changes in variation through a RR on Legendre polynomials of

*AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

Table 1. Characteristics of data and univariate random regression models fitted

Trait	Hereford								Wokalup							
	n^a	nc^b	$\bar{x}^c$	sd^d	p^e	G[f]	PE[f]	R[g]	n	nc	$\bar{x}$	sd	p	G	PE	R
JUN	2679	729	436	82	12	1	4/3	2	2889	808	463	85	9	1	3/3	2
JAN	2490	692	559	78	7	1	3/2	1	2625	722	592	84	5	1	3/1	1
GAIN	2398	663	128	41	3	1	1	1	2468	702	132	47	3	1	1	1
LOSS	2504	715	-78	48	6	1	3/1	2	2661	774	-89	53	4	1	1	2
WW	1985	631	244	44	4	1	1	1	2203	777	269	45	4	1	1	1

[a] No. of records [b] No. of cows [c] Mean (kg) [d] Standard deviation (kg) [e] No. of parameters [f] Order of fit/Rank fitted: G genetic, PE permanent environmental [g] No. of residual classes

age. For WW, the random effect of the calf's sire was also included (intercept only). Estimates of covariance components were obtained via restricted maximum likelihood (REML) using WOMBAT (Meyer 2007) with a sampling approach (Meyer and Houle 2013) to approximate standard errors.

Numerous univariate RR analyses were carried out for each trait, considering different orders of polynomial fit (up to quartic), as well as reduced rank estimates of the covariance matrices among RR coefficients, fitting separate measurement error variances for each year of age. The most parsimonious model was then selected based on the REML form of the Bayesian information criterion (BIC), and additional analyses decreased the number of error variances as far as possible without increasing BIC. The final model fitted 2 error variances (2, 3-8 years) for LOSS and JUN and a single error variance otherwise (see Table 1). Bi- and trivariate RR analyses were performed fitting the best model thus identified for each trait, again reducing rank of fit if eigenvalues close to zero were encountered.

RESULTS AND DISCUSSION

As illustrated in Figure 1, there was substantial variation in weight changes between cows. Cows continued to grow till 4 or 5 years of age and body weight changes depended on cow size, resulting in lower LOSS and somewhat higher GAIN at younger ages. Stringent model selection meant that a single coefficient was fitted for genetic effects for all traits, i.e. genetic covariances were considered constant for all ages. Quadratic or cubic polynomials were required to model changes in variation with age for permanent environmental effects on individual weights. For weight changes, however, these higher order effects mostly cancelled out, so that a simple repeatability model appeared appropriate for GAIN in both breeds. For LOSS, differences were most pronounced for heifers and required separate measurement error variances for heifers and older cows (see Table 1).

With calving spread over about three months, month of calving had a strong effect on LOSS and GAIN which was similar for both breeds. Figure 2 gives estimates for their effects, scaled to sum to zero. As expected, cows not raising a calf (month 0) had substantially larger gains and lost less weight. LOSS and, in turn, GAIN were least for the small proportion of cows calving late (month=6) as calving or depletion of body reserves in early lactation occurred after the June weighing.

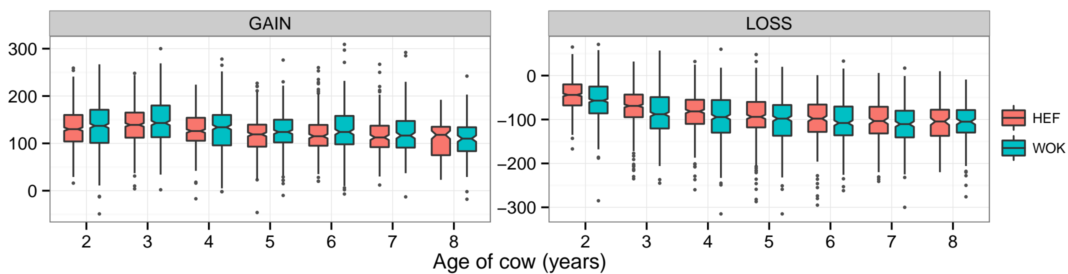


Figure 1. Distribution of cow weight changes (kg).

Estimates of phenotypic variances and corresponding variance ratios from trivariate analyses are summarized in Figure 3 and Figure 4, respectively, with vertical bars showing the range of plus or minus one standard deviation. Genetic correlations between cow weights at different ages are generally found to be high. Hence, not surprisingly, heritability estimates for weight changes were low. Rose *et al.* (2013) reported somewhat higher estimates for weight changes, especially GAIN, of Merino ewes in Western Australia. Analogous arguments held for PE effects of cows, and corresponding repeatabilities were thus also low, 15 to 17% for GAIN and 8 to 26% for LOSS. Fitting a quadratic regression for PE effects for LOSS in HEF resulted in a corresponding shape for estimates of the pertaining variances which may reflect so-called 'end-of-range' problems often encountered in RR analyses. Treating WW as trait of the cow, variation between animals reflects a cow's potential for growth transmitted to the calf as well as her maternal effects. Estimates of variance ratios were consistent with results from previous analyses of WW as trait of the calf (Meyer *et al.* 1993), which identified maternal effects for HEF to be twice as important as for WOK.

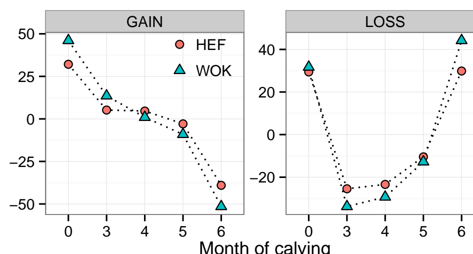


Figure 2. Month of calving effects (kg).

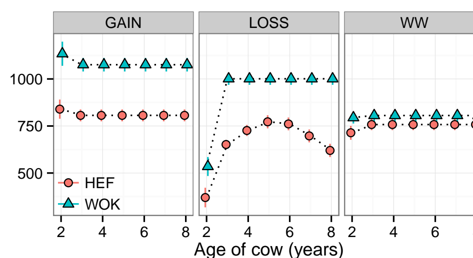


Figure 3. Phenotypic variances (kg^2).

Corresponding estimates of correlations are summarised in Figure 5 and selected results from bivariate analyses are given in Table 2. Genetic (r_G) and permanent environmental correlations between GAIN and LOSS were essentially unity, i.e. cows pre-disposed to lose more weight were also likely to gain more subsequently. However, with most variation due to environmental effects, corresponding phenotypic correlations were weak. Estimates of r_G between individual weights and weight changes ranged from 0.6 to 0.9 (absolute value; Table 2), emphasizing that genetically larger cows had the genetic potential for larger weight changes.

While phenotypic associations between seasonal changes and WW lacked strength, there were low to moderate genetic and permanent environmental correlations indicating that cows with more seasonal weight changes weaned larger calves. Of course, this was largely explicable by the genetic association between potential for growth of the cow and her calf. Correlations between GAIN and WW were somewhat weaker for HEF than for WOK. While differences were well within the range of sampling variation, this may reflect some dissimilarity in maternal capability, especially milk production. WOK are a synthetic breed comprising 25% Friesians, so that milk production is not considered a limiting factor. In contrast, estimates of maternal PE effects on WW in Hereford are consistently much higher, around 20%, than in most other breeds. Anecdotally this is often attributed

Table 2. Estimates of correlations from selected bivariate analyses

Traits	HEF				WOK			
	JUN GAIN	JAN LOSS	JUN WW	JAN WW	JUN GAIN	JAN LOSS	JUN WW	JAN WW
Genetic	0.88	-0.86	0.80	0.58	0.57	-0.69	0.91	0.89
Permanent environment[a]	0.46	-0.71	-0.75	-0.65	0.70	-0.94	0.16	-0.43
Phenotypic[a]	-0.13	-0.16	-0.01	-0.01	-0.13	-0.09	0.25	0.24

[a] Correlation for cows at 4 years of age

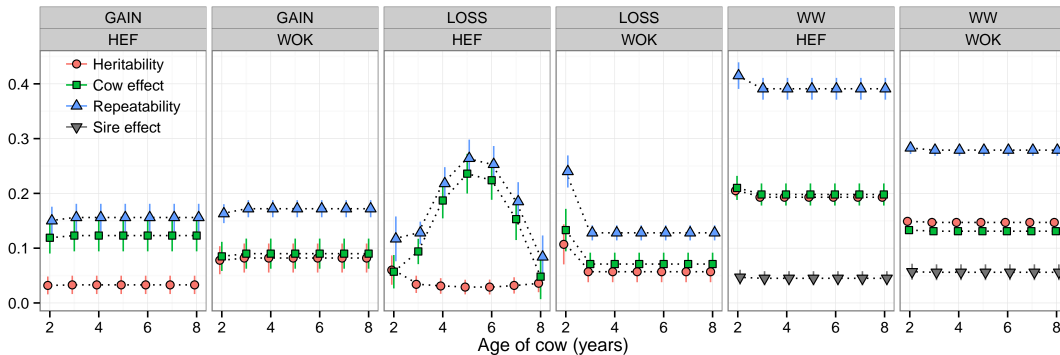


Figure 4. Estimates of variance ratios from trivariate analyses.

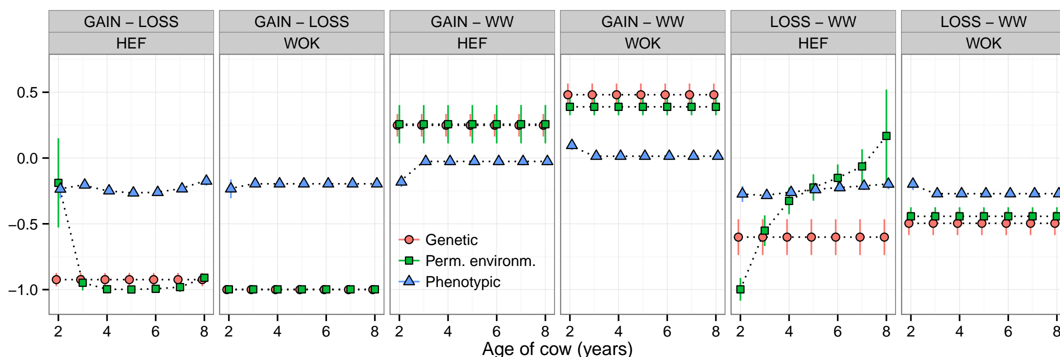


Figure 5. Estimates of correlations from trivariate analyses.

to poor milk production or short lactations.

CONCLUSIONS

Beef cows in pasture based production systems are likely to show strong seasonal fluctuations in body weight, with substantial variation between animals. However, most variation appears to be due to environmental effects, with low heritabilities and repeatabilities for weight changes. Estimates of genetic correlations suggest that larger cows are likely to lose and subsequently gain more weight. Selecting for robustness relies on proximate measures such as body weight change to predict through genetic parameters the capacity of the animal to achieve production goals in the face of environmental challenges while maintaining its ability to express functional traits. Current results suggest there may be limited scope to enhance robustness by maintaining body weight reserves of the cow while selecting for calf weaning weights in a pasture based Mediterranean production environment.

ACKNOWLEDGEMENTS

Work was supported by Meat and Livestock Australia grant B.BFG.0050

REFERENCES

Meyer K. (2007) *J. Zhejiang Uni. SCIENCE B* **8**:815.

Meyer K., Carrick M.J. and Donnelly B.J.P. (1993) *J. Anim. Sci.* **71**:2614.

Meyer K. and Houle D. (2013) *Proc. Ass. Advan. Anim. Breed. Genet.* **20**:523.

Rose G., Kause A., Mulder H.A., van der Werf J.H.J., Thompson A.N., Ferguson M.B. and van Arendonk J.A.M. (2013) *J. Anim. Sci.* **91**:2555.

Proc. Assoc. Advmt. Breed. Genet. 21: 197-200

EFFECTS OF PRODUCTION LEVEL ON THE EXPRESSION OF BREED AND HETEROSIS FOR LACTATION YIELDS OF MILK, FAT AND PROTEIN IN COWS MILKED ONCE- AND TWICE-DAILY

F. Lembeye[1], N. Lopez-Villalobos[1], J.L. Burke[1] and S.R. Davis[2]

[1]Institute of Veterinary, Animal and Biomedical science, Massey University, Palmerston North, New Zealand
[2]Livestock Improvement Corporation, Hamilton, New Zealand

SUMMARY

The objective of this study was to estimate breed and heterosis effects between Holstein-Friesian (F) and Jersey (J) cows at different herd production levels in herds milked once-a-day (OAD) or twice-a-day (TAD) in New Zealand. Three groups of herd production levels based on milksolids (MS, fat + protein) production per cow were considered: low, medium and high. The average MS production per cow was: 203.6; 269.2 and 339.9 kg of MS in herds milked OAD; and 272.7; 353.8 and 434.1 kg of MS in herds milked TAD. Data consisted of 322,327 lactation records from 35,192 F; 31,118 J and 88,606 crossbred (F×J) cows that calved in spring between 2008 and 2012. Breed effects, defined as F-J, increased as production level of the herd increased in both systems. Heterosis effects, expressed as a percentage of the mean of the parental breeds, ranged between 3.3 and 8.4% in OAD and 4.4 and 7.4% in TAD systems. The highest expressions of heterosis were found at medium (6.3-8.4%) and high (6.1-7.4%) production level in cows milked OAD and TAD, respectively. In conclusion, production level affects the expression of breed and heterosis in both milking systems. Breed and heterosis effects increased as production levels increased.

INTRODUCTION

Pasture-based dairy farming in New Zealand has predominantly been with cows milked twice-a-day (TAD). However, since the late 1990s, milking once-a-day (OAD) has been adopted by some farmers for herd management and lifestyle benefits (Davis, 2005).

Crossbreeding in New Zealand has brought favourable heterosis for production, fertility and survival traits, which results in increased farm profitability (Lopez-Villalobos *et al.* 2000). In an extensive review, Barlow (1981) concluded that heterosis was better expressed when the environmental conditions are sub-optimal, but in New Zealand, Bryant *et al.* (2007) found low or no heterosis on restricted environments in TAD systems.

Because there is evidence of different breed performances and expression of heterosis in different environments in cows milked TAD (Bryant *et al.* 2007; Penasa *et al.* 2010; Kargo *et al.* 2012), the objective of this study was to estimate breed and heterosis effects at different production levels (as an indication of dry matter intake) in cows milked OAD and compare the results with cows milked TAD under New Zealand conditions.

MATERIALS AND METHODS

Data. Lactation yields of milk (MY), fat (FY) and protein (PY) were provided by Livestock Improvement Corporation for the period 2008-2012. Initial data was restricted as follows. Lactation records were sorted based on a code to determine if the cow was milked OAD or TAD at a specific lactation record. In the present study, OAD herds were considered as those herds in which 100% of the cows were milked OAD during the entire lactation. Twice-a-day herds were selected in a radius of 20 km from OAD herds using map coordinates. In some cases, in a given single map co-ordinate, OAD herd was surrounded by several TAD herds, in that case all TAD

herds were selected using the GPS Visualizer (Schneider, 2012). Only herds with more than 50 cows recorded per season were used in the analysis. Only records from spring calving cows in their first five lactations with lactation lengths greater than 150 days and less than 305 days were considered. Also, only records from Holstein-Friesian (F), Jersey (J) and their crosses (F×J) were kept, discarding cows whose parents provided no information about their breed composition.

After all the restrictions were imposed, the dataset contained 322,327 lactation records from 154,916 cows (35,192 F; 88,606 F×J and 31,118 J); 127,885 lactations were from 298 herds milked OAD and 194,442 lactations were from 350 herds milked TAD.

Three groups (clusters) per milking frequency were constructed based on herd production levels (low, medium or high) for milksolids (MS, fat + protein) per cow using the FASTCLUS procedure of SAS version 9.3 (SAS Institute Inc., Cary, NC., USA, 2012). Low, medium and high production levels were considered those herds which respectively yielded: 203.6; 269.2 and 339.9 kg of MS in herds milked OAD; and 272.7; 353.8 and 434.1 kg of MS in herds milked TAD. Number of herds per each cluster was: 110, 141 and 47 for low, medium and high production levels in the OAD population; and 168, 150 and 32 in the TAD population.

Statistical Analysis. A univariate linear model was used to obtain breed and heterosis effects for MY, PY and PY using the MIXED procedure (SAS 2012). The model included the random effect of herd-season, the fixed effects of milking frequency, lactation number, production level, interaction between milking frequency and lactation number, interaction between milking frequency and production level, linear regression of MY, FY or PY on mean calving date deviation from median calving date of the herd for a given season, linear regressions of MY, FY or PY on proportion of F within each combination between production levels and milking frequencies, linear regressions of MY, FY or PY on coefficient of heterosis within each combination of production level and milking frequencies and the random residual error. The solutions for fixed effects and estimates of the regression coefficients for proportion of F and heterosis coefficients were used to predict the performance of F, J and F_1 F×J cows at different production levels.

RESULTS AND DISCUSSION

Table 1 presents predicted production level of pure F, J and crossbred F_1 (F×J), with breed and heterosis effects for MY, FY and PY in each combination of milking frequency and production level.

Breed effects, defined as F-J, increased as production levels of the herd increased in both OAD and TAD systems. The superiority of F cows at high production levels showed more than double the level observed for yield of milk, fat and protein at low and medium production level in both systems. The smaller breed effect at low and medium production level compared to high production level suggest that J cows might have an advantage over F cows in those environments, in particular in OAD systems. The nutritional status of cows in grazing conditions varies considerably across the seasons in New Zealand; hence F cows cannot express their potential when they are exposed to restrictive periods (Ahlborn-Breier and Hohenboken, 1991).

Differences in productive performance among breeds relate to the environment in which the breeds are evaluated (Bryant *et al.* 2007; Penasa *et al.* 2010; Kargo *et al.* 2012). Those studies reported that in general, more productive cows (with large proportion of North American genes) increased their superiority for MY in higher input systems. In more intensive systems, the nutritional requirements of high productive cows are likely better achieved (Penasa *et al.* 2010) allowing high producing cows (as F cows) to express their genetic merit for milk, fat and protein production.

Table 1. Breed performance and standard errors of production traits for Holstein-Friesian (F), Jersey (J) and first cross (F_1) F×J cows, and estimates of breed and heterosis effects at different production level

Production level[1]	MF[2]	F	F_1 F×J	J	Breed effect	Heterosis effect	
					F-J (kg)	kg†	%‡
			Milk yield (kg/cow)				
L	1	2572±26	2479±24	2101±25	471^a±18	143^a±15	6.1
	2	3526±18	3319±17	2760±19	767^b±12	176^a±10	5.6
M	1	3305±21	3193±20	2703±20	602^a±12	189^a±10	6.3
	2	4520±17	4198±17	3523±18	997^b±11	177^a±8	4.4
H	1	4221±29	3901±28	3331±29	890^a±15	125^a±13	3.3
	2	5595±26	5141±27	4096±29	1499^b±15	295^b±13	6.1
			Fat yield (kg/cow)				
L	1	121.4±1.3	132.7±1.1	118.6±1.2	2.8^a±0.9	9.3^a±0.7	7.7
	2	157.7±0.8	178.9±0.8	151.7±0.9	6.0^b±0.6	11.4^a±0.5	7.3
M	1	157.2±1.0	178.9±0.9	155.9±0.9	1.3^a±0.6	11.1^a±0.5	7.1
	2	202.0±0.8	231.4±0.8	195.9±0.9	6.1^b±0.5	13.5^b±0.4	6.8
H	1	195.8±1.4	209.4±1.3	191.3±1.4	4.5^a±0.7	10.0^a±0.6	5.2
	2	245.5±1.2	267.4±1.3	233.4±1.4	12.2^b±0.7	17.8^b±0.6	7.4
			Protein yield (kg/cow)				
L	1	97.7±1.0	99.3±0.9	87.6±1.0	10.1^a±0.6	6.7^a±0.5	6.7
	2	127.6±0.7	127.1±0.7	111.0±0.7	16.7^b±0.4	7.8^a±0.3	6.6
M	1	126.2±0.8	128.7±0.7	114.4±0.8	11.8^a±0.4	8.4^a±0.4	8.4
	2	164.1±0.6	162.4±0.6	142.8±0.7	21.3^b±0.4	9.0^a±0.3	5.9
H	1	160.1±1.1	157.5±1.1	141.2±1.1	18.9^a±0.5	6.9^a±0.5	6.9
	2	203.9±1.0	199.6±1.0	171.3±1.1	32.6^b±0.5	12.0^b±0.5	6.4

[1] L= low milksolids (fat + protein) yield, M = medium milksolids yield, H = high milksolids yield.
[2] MF = milking frequency, 1 = milking once-daily and 2 = milking twice-daily.
† Expressed as F_1 F×J – (F + J)/2.
‡ Expressed as a percentage of heterosis effects relative to the phenotypic average of the parental breeds under milking frequency and production levels, as appropriate.
[a,b] Within traits and production level, breed and heterosis effects with different superscripts were significantly different between milking frequencies (P<0.05).

Heterosis effects for production traits, expressed in absolute values, tended to be greater in TAD, but in relative values, heterosis effects were similar in both, OAD and TAD systems (3.3-8.4% in OAD and 4.4-7.4% in TAD systems).

The expression of heterosis for production traits was influenced by production levels. In TAD systems the absolute values of heterosis effects for milk, fat and protein increased as production level increased but in relative values, heterosis at low and high production levels tended to be similar. The lowest heterosis effects expressed in relative values were observed at medium production levels. In OAD systems, the absolute and relative heterosis effects for the production traits were greater at medium compared to low and high production levels.

Despite lower relative heterosis effects at low production level these effects are similar to the heterosis effects for production traits in New Zealand by Ahlborn-Breier and Hohenboken (1991) and Harris (1996).

The results obtained in this study are similar to the studies of Bryant *et al*. (2007) and Kargo *et al*. (2012), who found greater heterosis in the medium and high producing environments, contradicting Barlow (1981) who affirmed that heterosis effects tended to be greater in less supportive environments.

CONCLUSION

Expression of breed and heterosis effects differed across milking frequencies and production levels. The productive performance of F cows relative to J cows increased as production levels of MS increased in both, OAD and TAD systems. Production levels of the herds are also a factor which affects the expression of heterosis in both milking frequencies.

ACKNOWLEDGMENTS

The principal author acknowledges the support from Programa Formación de Capital Humano Avanzado, Becas Chile, CONICYT doctoral scholarship and information provided by Livestock Improvement Corporation (Hamilton, New Zealand).

REFERENCES

Ahlborn-Breier G. and Hohenboken W.D. (1991) *J. Dairy Sci*. **74**: 592.

Barlow R. (1981) *Anim. Breed. Abstr.* **49**: 715.

Bryant J.R., Lopez-Villalobos N., Pryce J.E., Holmes C.W., Johnson D.L. and Garrick D.J. (2007) *J. Dairy Sci.* **90**: 1548.

Davis S.R. (2005). *N.Z. Vet. J.* **53**: 400.

Harris B.L., Clark J.M. and Jackson R.G. (1996). *Proc N.Z. Soc. Anim. Prod.* **56**: 12.

Lopez-Villalobos N., Garrick D.J., Holmes C.W., Blair H.T. and Spelman R.J. (2000) *J. Dairy Sci.* **83**: 144.

Kargo M., Madsen P. and Norber E. (2012*) J. Dairy Sci*. **95**: 925.

Penasa M., De Marchi M., Dal Zotto R., de Jong G, Bittante G. and Cassandro M. (2010) *Livest. Sci*. **131**: 52.

Schneider A. (2012) About GPS Visualizer. http://www.gpsvisualizer.com/about.html (accessed 20 June 2014).

Proc. Assoc. Advmt. Breed. Genet. 21: 201-204

REVISITING TOTAL WEANING WEIGHT AS A SELECTION CRITERION

K.L. Bunter and D.J. Brown

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW, 2350

SUMMARY

Data from a single prolific Merino flock (N=10705 joining records) recorded over 10 years were used to estimate genetic parameters for annual total weight weaned (TWW) of ewes which weaned lambs (N=8615), treated as a ewe trait, and the accompanying reproductive traits. TWW was the sum of individual weights of weaned lambs to birth ewe. An alternative trait definition included ewes which lambed and lost, which received zero trait values (TWW0, N=9509). Both TWW and TWW0 were lowly heritable (range h^2: 0.06-0.11). Most of the phenotypic variation in these traits resulted from variation in the number of lambs weaned. Trait definition significantly influenced both the observable variation in the ewe weaning weight traits (eg. TWW vs TWW0) and correlations with reproductive traits. Because total weight weaned traits combine direct and maternal effects, and multiple non-genetic sources of variation, prediction of response to selection for total weight weaned and its components depends on the trait definition used and accompanying population characteristics and genetic parameters. We conclude that selection on an index which combines breeding values for reproductive performance, and both direct and maternal contributions to weaning weight traits, should be considered to improve ewe productivity in a more predictable manner under dual purpose breeding goals. This index is provided by Sheep Genetics, which also appropriately analyses individual animal reproductive and weight data while accounting for systematic effects and multiple records.

INTRODUCTION

Total lamb weight weaned per ewe joined has been proposed as a simple selection criterion for increasing reproduction and ewe productivity in dual purpose sheep (Snowder and Fogarty 2009). Total lamb weight weaned can reflect the full complement (or a subset) of traits important to ewe productivity, such as conception, ewe survival, litter size and lamb survival, along with the ewe's maternal contributions (genetic and non-genetic) to lamb weaning weight(s). However, lamb weaning weights are also influenced by genes of the lamb (the direct genetic effect), half of which were received from the sire. Individual lamb weaning weights are also significantly influenced by a number of non-genetic factors, such as season, age of dam, birth and rear type, lamb gender, and weaning age (Ch'ang and Rae 1961). Therefore, trait values for total weight weaned combine many sources of variation, several of which are non-genetic in origin. In this study we estimated parameters for weaning weight traits defined as traits of the ewe in a prolific Merino population, particularly with respect to illustrating the effect of using alternative trait definitions and correction for non-genetic effects, to investigate potential implications of using a complex selection criterion such as total lamb weight weaned.

MATERIALS AND METHODS

Data were obtained from a prolific (high fertility, high litter size) Merino population recorded over 10 years for reproductive performance, as described in Bunter *et al.* (2014). Ewes with reproductive records (N=7457) were daughters of 308 sires and 3540 dams. A subset of individual lamb weaning weights was obtained over 8 years. Lambs recorded with weaning weights were progeny of 4197 ewes and 136 service sires.

*AGBU is a joint venture of NSW DPI and the University of New England

Trait definitions. Reproductive traits for ewes included fertility (FERT), the number of lambs born (NLB) and weaned (NLW) per ewe joined, along with litter size at birth (LSIZE) and at weaning (LSIZEW) for lambed ewes. Weaning weight traits were defined as the average (AVWW) and total weaning weight (TWW) calculated annually for each ewe from weaning weight records on their lambs. For comparison, ewes which lambed but failed to wean a lamb had records augmented with trait values of zero (AVWW0, TWW0). Infertile ewes which did not lamb within a year had (zero) records for FERT, NLB and NLW only.

Models for analyses. All traits were treated as repeated measures of the ewe under an animal model, with additional variation due to service sire effects (σ^2s). The additive direct genetic contribution to lamb weaning weight was approximated as $4\sigma^2$s/σ^2p using parameters for AVWW. Systematic effects for ewe reproductive traits included year (10 levels) combined with lambing contemporary group (CGP: 30 levels), with CGP defining conception method (AI vs natural), ewe age group (3 levels) and breeder defined management groups. Models for weaning weight traits included year (8 levels) and weaning CGP (58 levels), where CGP included ewe age group and breeder defined management groups for weaning traits. Parameters were first estimated in univariate analyses using models without any covariates. Litter size at lambing, the number of lambs weaned and weaning age were then added as linear covariates for weaning weight traits (AVWW and TWW) for comparison. For zero augmented traits (eg TWW0 and AVWW0) no weaning contemporary group was defined and the covariate for lamb age at weaning was excluded from models for analyses. The relative contributions of each covariate to phenotypes for AVWW and TWW were approximated as the squared correlation between the weaning weight trait and each covariate, calculated as $(b*SD_X/SD_Y)$^2, where b is the partial regression coefficient, SD_X and SD_Y are the SD of each covariate (X) and the dependent trait (Y), with both X and Y pre-adjusted for year-CGP effects. Correlations between specific traits were estimated from a series of bivariate analyses using ASReml (Gilmour *et al.* 2006), excluding all covariates from models for analyses.

RESULTS AND DISCUSSION

Ewe average weaning weight of lambs at approximately 107 days of age was 25.6 kg (Table 1). Total weaning weight averaged 34.3 kg and was highly variable (CV=32%) relative to AVWW (CV=16%). Mean values decreased, while phenotypic variance and CV for both traits increased when the data were augmented for ewes which lambed and lost (ie 0 kg weight weaned).

Parameter estimates. Heritability estimates were low and close to expectation for ewe reproductive traits. Variation due to service sire (σ^2s) was significant for FERT but not litter size. Direct heritability for lamb weaning weight (as calculated from σ^2s) was moderate regardless of litter size: 0.23±0.04 estimated for single born lambs versus 0.21±0.04 across all lambs weaned. Direct heritability was lower than the 0.29±0.01 reported by Safari *et al.* (2007) from a more diverse Merino population. Variance due to the permanent environmental effect of the ewe was similar across these studies (0.04, derived from values in Table 1, vs 0.05). In our analyses, when service sire was not fitted in models for analyses, service sire variance was mostly repartitioned to the residual variance (not presented).

Parameter estimates for AVWW (h^2=0.10±0.02, pe^2=0.04) were relatively low. The expectation for component(s) contributing to the calculated ewe AVWW is $1/4\sigma^2a+\sigma^2m+\sigma am+\sigma^2c$, where: σ^2a is the additive genetic variance (direct effect), σ^2m is the additive maternal genetic variance, σam is the direct-maternal covariance, and σ^2c is the common litter effect. Using parameters for individual lamb weaning weights estimated by Safari *et al.* (2007), assuming σam=0 and accurate partitioning for σ^2c, the heritability for AVWW could be approximated as 0.25×0.26+0.10=0.17, which is higher than the value of 0.10±0.02 obtained here. Compared to phenotypic variance of individual lamb weaning weights, the phenotypic variance of AVWW is reduced. Heritability and repeatability for TWW were similar to estimates for AVWW,

but the phenotypic variance was approximately doubled. Each TWW record is equivalent to AVWW×*n*, where *n* was the number of progeny recorded at weaning, but additional variation is also expected due to ewe genetic contributions to *n* and the (co)variances between reproductive and weaning weight traits. Relative to AVWW and TWW, phenotypic variances were greatly increased by the zero enrichment of AVWW0 and TWW0. However, contemporary groups and a covariate for weaning age cannot sensibly be applied across values for these trait definitions.

Table 1. Raw data characteristics, along with estimates of heritability (h^2), repeatability (r), and service sire (σ^2s), residual (σ^2e) and phenotypic (σ^2p) variances from univariate analyses (*line 1: no covariates; line 2: covariates included, with % reduction of variance in brackets)

Trait	N	Mean (SD)	h^2	r	σ^2s	σ^2e	σ^2p
FERT	10705	0.95 (0.22)	0.02±0.01	0.03±0.01	0.05±0.01	0.042	0.046
NLB	10705	1.58 (0.69)	0.07±0.01	0.13±0.01	0.02±0.01	0.381	0.450
NLW	10705	1.18 (0.66)	0.03±0.01	0.09±0.01	0.02±0.00	0.361	0.406
LSIZE	10139	1.66 (0.60)	0.10±0.02	0.15±0.01	0.01±0.00	0.289	0.347
LSIZEW	10139	1.28 (0.63)	0.03±0.01	0.10±0.01	0.01±0.00	0.334	0.374
AVWW	8615	25.6 (4.20)	0.08±0.02	0.13±0.01	0.05±0.01	14.3	17.4
			0.10±0.02	0.14±0.01	0.05±0.01	9.55 (34)	11.9 (32)
AVWW0	9509	23.2 (8.47)	0.06±0.01	0.14±0.02	0.02±0.00	58.5	69.7
			0.04±0.01	0.07±0.01	0.02±0.00	44.5 (24)	49.0 (30)
TWW	8615	34.3 (11.0)	0.06±0.01	0.11±0.01	0.04±0.01	101	119
			0.11±0.02	0.12±0.01	0.05±0.01	19.9 (80)	25.4 (79)
TWW0	9509	31.1 (14.5)	0.05±0.01	0.12±0.01	0.02±0.00	173	199
			0.08±0.01	0.12±0.01	0.07±0.01	20.3 (88)	24.9 (87)

$h^2=\sigma^2a/\sigma^2p$ and $r=(\sigma^2a+\sigma^2pe)/\sigma^2p$, where σ^2a is the additive genetic variance and σ^2pe is variance due to permanent environmental effects of the ewe; *covariates relevant for ewe weaning weight traits only

The importance of weaning age, litter size and lambs weaned. Models without covariates explained <5% of the variation in all ewe weaning weight traits. When all covariates were included in the models for analyses, phenotypic variances were greatly reduced: by 30-32% for AVWW and AVWW0 and by 79-87% for TWW and TWW0 (Table 1). Variation in weaning age and litter size at birth explained about 13-15% each of the variation in AVWW. The number of lambs weaned explained the bulk of variation in TWW (r^2~82%) (results not tabulated). Birth-rearing class and weaning age are the main factors affecting individual weaning weights of lambs (Ch'ang and Rae 1961) and consequently traits derived from lamb weights for their dams. Since weaning dates are generally fixed, variation in weaning age mostly resulted from how early ewes conceived in the joining period. In these data, heritability from an additional analysis for the number of days until lambing, after the commencement of lambing, was only 0.03±0.01 (r=0.05±0.01). Therefore, for accurate comparisons amongst ewes, ewe weaning weight traits should also be corrected for lamb age at weaning.

Correlations between reproductive and weaning weight traits. Correlations between traits at the genetic and phenotypic level indicate that fertility is favourably correlated with all weaning weight traits (Table 2). Both TWW and TWW0 also had consistently positive correlations with reproductive traits (NLB, NLW, LSIZE and LSIZEW), being larger in magnitude for reproductive traits representing lambs alive at weaning. This is partly because only weaned lambs generate non-zero weaning weight records. In contrast, some unfavourable correlations were evident between the reproductive traits and AVWW or AVWW0, demonstrating that individual lamb weights are decreased for lambs weaned in larger litters. Genetic correlations between AVWW and TWW or AVWW0 and TWW0 were 0.61±0.04 and 0.80±0.07 (not tabulated). These results suggest overall

that selection for (unadjusted) TWW would most strongly favour litter size at weaning in this prolific Merino flock, thereby increasing total weight weaned, but individual weaning weights would suffer. The latter has implications for lamb marketability at weaning and/or post-weaning survival of lambs.

Table 2. Additive genetic (ra) and phenotypic (rp) correlations between reproductive and weight traits

Trait	Correlation	AVWW	AVWW0	TWW	TWW0
FERT*	ra	0.09±0.17	0.81±0.21	0.10±0.17	0.22±0.18
NLB	ra	-0.20±0.14	-0.32±0.13	0.51±0.11	0.18±0.14
	rp	-0.38±0.01	-0.18±0.01	0.53±0.00	0.36±0.00
NLW	ra	0.12±0.21	0.68±0.12	0.89±0.04	0.92±0.03
	rp	-0.52±0.01	0.42±0.01	0.93±0.00	0.94±0.00
LSIZE	ra	-0.22±0.12	-0.45±0.11	0.51±0.10	0.09±0.14
	rp	-0.34±0.01	-0.21±0.06	0.48±0.01	0.31±0.00
LSIZEW	ra	0.23±0.24	0.57±0.15	0.88±0.06	0.80±0.07
	rp	-0.53±0.07	0.39±0.01	0.87±0.00	0.84±0.00

*residual and therefore phenotypic correlations are not estimable

While simple in concept, total weaning weight is an exceptionally complex trait. Trait values for ewes represent both direct and maternal effects, correlations between traits across trait groups, non-genetic factors influencing both ewe and lamb performance, the possibility of unaccounted for environmental covariance between dam and offspring, and a potentially high degree of variance inflation due to inclusion of zero values, which is somewhat arbitrary. This can create quite large fluctuations in genetic parameters (eg see correlations between LSIZE and TWW or TWW0, which are affected by lamb survival). Therefore, choice of trait definition and the ability to adjust for systematic effects will impact on the expected response to selection for total weaning weight, and correlated response in the sub-traits of economic importance. Studies to date have typically not made these calculations.

CONCLUSION

Selection for total weaning weight is simple at face value, but the response to selection for contributing traits will vary depending on population characteristics, the trait definition used, the corrections for non-genetic effects applied and therefore underlying genetic parameters. Further work is required to evaluate whether index selection combining ewe reproductive traits with both direct and maternal components for weaning weight, as is included in the existing Sheep Genetics dual purpose Merino index, delivers a more optimal and predictable improvement in response to selection for ewe reproductive traits and productivity, and individual lamb weaning weights, when compared to selection based on TWW alone.

REFERENCES

Bunter K.L., Swan A.A., Purvis I.W., Brown D.J. (2015) *Anim. Prod. Sci.* (accepted November 3, 2014, http://dx.doi.org/10.1071/AN14630).

Ch'ang T.S. and Rae A.L. (1961) *NZ J. Agric. Res.* **4**: 578.

Gilmour A.R., Gogel B.J., Cullis B.R., Thompson R. (2006) 'ASREML User Guide 3.0.' (VSN International)

Safari E., Fogarty N.M., Gilmour A.R., Atkins K.D., Mortimer S.I., Swan A.A., Brien F.D., Greeff J.C., van der Werf, J.H.J. (2007) *Aust. J. Agric. Res.* **58:** 177.

Snowder G.D. and Fogarty N.M. (2009) *Anim. Prod. Sci.* **49:** 9.

Proc. Assoc. Advmt. Breed. Genet. 21: 205-208

IS FIBRE COMFORT FACTOR REQUIRED IN MERINO BREEDING PROGRAMS?

S. Hatcher[1,2] and D.J. Brown[1,3]

[1]CRC for Sheep Industry Innovation, Homestead Building, UNE, Armidale, NSW, 2351
[2]NSW DPI, Orange Agricultural Institute, Forest Road, Orange, NSW, 2800
[3]Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW, 2351

SUMMARY

Genetic parameters for fibre comfort factor (FCF) and correlations with key production traits were estimated from research (Cooperative Research Centre for Sheep Industry Innovation Information Nucleus) and industry data (Sheep Genetics MERINOSELECT). FCF is moderately to highly heritable and genetically consistent through life, with the yearling, hogget and adult expressions phenotypically (0.54 to 0.73) and genetically (0.97 to 0.98) correlated. The strong genetic correlations (r_g) between FCF and both fibre diameter (FD) and FD standard deviation (FDSD) (range -0.58 and -0.92) indicate that selection to reduce FD or FDSD will generate favourable correlated increases in FCF. There would be little to be gained from including FCF in Merino breeding programs with an existing emphasis on reducing FD and it would be difficult for medium to strong wool breeders to add FCF to their breeding programs and maintain FD.

INTRODUCTION

Coarser fibres in a fibre diameter distribution are responsible for the fabric-evoked prickle sensation felt by wearers of next-to-skin garments. Coarser fibre ends buckle less readily when fabric pushes against the skin during wear and mechanically stimulate particular nerve cells lying close to the skin surface (Naylor 1992). The threshold value of buckling force required to trigger the nerve cell response corresponds to a FD of approximately 30 µm (Naylor 1992) and this finding led to the use of FCF, the percentage of fibres in the FD distribution < 30 µm, as a means of categorising apparel fibres in terms of their propensity to cause prickle when worn next-to-skin. The prickle sensation is not exclusive to wool and depends on a large number of other parameters including fabric construction and physiological state of the wearer (Naylor 1992). Despite this, consumers in key markets for Australian wool consistently associate prickle with wool and many Merino breeders are seeking to genetically increase FCF to a level, >95%, beyond which prickle cannot be perceived by most people under normal conditions (Garnsworthy *et al.* 1988). While fine wool sheep will typically have high FCF levels due to their low average FD (Baxter and Cottle 1998), some breeders of medium and strong wool sheep have expressed interest in breeding sheep with a higher FCF without changing average FD. This paper reports the genetic parameters for FCF and the phenotypic and genetic correlations with liveweight, wool production and a suite of measured and visual wool quality traits in Merino sheep using a combination of industry and research data.

MATERIALS AND METHODS

Research data. Data from Merino progeny (n = 4,958) born into the Sheep CRC Information Nucleus (IN) Flock (van der Werf *et al.* 2010) between 2007 and 2010 were used. Yearling (Y, 10 – 13 months) and adult (A, 22 – 25 months) performance for a suite of wool production and quality traits were assessed or measured prior to the yearling and adult shearings. Fleece rot (FLROT), colour (COL), character (CHAR), dust penetration (DUST), staple weathering

* AGBU is a joint institute of NSW Department of Primary Industries and The University of New England

(WEATH) and staple structure (SSTRC) were scored using the Visual Sheep Scores guide (AWI Ltd and MLA Ltd 2013), with handle (HAND) scored according to Casey and Cousins (2010) and coverage (COV), fleece density (DENS) and nourishment (NOUR) assessed according to AMSEA guidelines (Casey *et al.* 2009). Right midside samples (approx. 80g) from each animal were measured at AWTA Limited (Melbourne) using standard IWTO test methods for yield (YIELD), staple length (SL), staple strength (SS), FD, FDSD, FD coefficient of variation (FDCV), FCF, mean fibre curvature (CURVE), brightness (Y), clean colour (Y–Z) and resistance to compression (RTOC). The unskirted greasy fleece weight (GFW) (belly wool included) was recorded at shearing with clean fleece weight (CFW) calculated as the product of GFW and YIELD. Both GFW and CFW were corrected to 365-day growth equivalents. Following shearing, the liveweight (LWT) of every sheep was recorded after being held off feed for approximately 2 hours.

Industry data. Pedigree and performance data were extracted from the Sheep Genetics (SG) MERINOSELECT database (Brown *et al.* 2007). A subset of flocks was selected based on their recording of FCF and were a mix of industry ram breeding, research and sire evaluation flocks. Only those animals with both sire and dam pedigree, and born from 2005 onwards were included. The traits analysed were LWT, GFW, CFW, FD, FDCV, FCF, CURVE, SL, SS and COL recorded at yearling and hogget (H, 13 – 18 months) ages. The pedigree was built using all ancestral information available.

ASReml 3.0 (Gilmour *et al.* 2009) was used to estimate variance components and genetic parameters. For both data sets, birth type, rearing type, and age of dam were fitted as fixed effects. Age of shearing was fitted as a covariate to the yearling IN data but fitted as fixed to the SG data. Flock and drop were also fitted as fixed to the IN data with genetic group, sire by flock and a maternal genetic effect fitted as random effects. For the SG data, a fixed effect of contemporary group (defined as flock, year of birth, sex, date of measurement, management group subclass), was also fitted for all traits along with random terms for the direct genetic effects and sire by flock year interaction. Maternal permanent environment effects were included for the SG LWT, GFW and CFW data with genetic groups, allocated on a flock basis for link flocks with sufficient data, fitted for all traits. Phenotypic and genetic correlations, with standard errors, were estimated from the appropriate covariances using a series of bivariate analyses.

RESULTS AND DISCUSSION

The mean FCF were similar for both data sources, as were the phenotypic variances which tended to increase with age (Table 1). Within the IN data, there was no evidence of sire by flock or maternal genetic effects for YFCF but these were both significant for AFCF representing 2.2 and 20.1% of the phenotypic variance respectively. For the SG data both the sire by flock and maternal genetic effects were also significant for YFCF (6% of the phenotypic variance) but not for HFCF.

Table 1. Mean, variance components, coefficient of variation (%) and heritability for FCF measured as yearlings, hoggets and adults from each data set

Trait	Mean	Variance components						CV	Heritability
		Phenotypic	Residual	Additive	Sire by flock	Maternal	Genetic gp	(%)	(h^2)
IN									
YFCF	99.58	0.34±0.01	0.25±0.01	0.09±0.01	-	-	0.07±0.06	0.59	0.27 ± 0.04
AFCF	99.40	1.34±0.03	0.72±0.04	0.32±0.04	0.03±0.02	0.27±0.06	0.20±0.18	1.16	0.20 ± 0.04
SG									
YFCF	99.60	0.47±0.01	0.35±0.01	0.07±0.01	0.03±0.00	0.03±0.00		0.63	0.14 ± 0.01
HFCF	99.11	1.52±0.02	0.22±0.02	1.30±0.03	0.01±0.00	0.00±0.00		1.82	0.85 ± 0.03

Genetic group was a significant source of variation in both the IN yearling and adult FCF,

representing 21 and 15% respectively of the phenotypic variance, however this is not surprising given the genetic grouping of the IN Merino progeny is largely based on a FD classification via their pedigree (i.e. ultra/superfine, fine fine/medium, medium strong) and the strong relationship between FCF and FD (Baxter and Cottle 1998). The heritability estimates for the IN yearling and adult FCF were both moderate, however the SG YFCF estimate was low (approximately half that of IN) while that for HFCF was high due to low residual and high additive variance at the hogget expression.

Table 2: Phenotypic and genetic correlations (± s.e.) between a) yearling and adult FCF and yearling LWT, wool production, wool quality and visual wool quality scores estimated from the IN and b) yearling and hogget FCF and yearling and hogget LWT, wool production, wool quality and greasy colour from SG.

a)

Trait*	Yearling FCF		Adult FCF	
	r_p	r_g	r_p	r_g
YLWT	-0.09 ± 0.02	-0.25 ± 0.09	-0.13 ± 0.03	-0.22 ± 0.07
YGFW	-0.15 ± 0.02	-0.37 ± 0.09	-0.05 ± 0.03	-0.05 ± 0.08
YCFW	-0.14 ± 0.02	-0.34 ± 0.09	-0.06 ± 0.03	-0.06 ± 0.08
YFD	-0.50 ± 0.02	-0.78 ± 0.05	-0.46 ± 0.02	-0.58 ± 0.05
YFDSD	-0.62 ± 0.01	-0.89 ± 0.04	-0.46 ± 0.02	-0.57 ± 0.06
YFDCV	-0.32 ± 0.02	-0.36 ± 0.08	-0.18 ± 0.03	-0.15 ± 0.07
YCURVE	-0.08 ± 0.02	-0.16 ± 0.09	-0.06 ± 0.02	-0.32 ± 0.16
YRTOC	-0.23 ± 0.02	-0.48 ± 0.08	-0.33 ± 0.03	-0.57 ± 0.06
Y(Y-Z)	-0.12 ± 0.02	-0.24 ± 0.09	-0.11 ± 0.03	-0.17 ± 0.08
YCOL	-0.12 ± 0.02	-0.40 ± 0.10	-0.17 ± 0.03	-0.43 ± 0.08
YCHAR	-0.20 ± 0.02	-0.71 ± 0.08	-0.27 ± 0.03	-0.73 ± 0.06
YSSTRUC	-0.27 ± 0.02	-0.86 ± 0.07	-0.29 ± 0.02	-0.77 ± 0.06
YHAND	-0.20 ± 0.02	-0.63 ± 0.11	-0.15 ± 0.03	-0.36 ± 0.11
YNOUR	-0.02 ± 0.02	-0.24 ± 0.14	-0.06 ± 0.03	-0.34 ± 0.13

*IN traits listed in the materials and methods with 4 negligible correlations (|r|<0.2) are omitted from the table.

b)

Trait	Yearling FCF		Trait	Hogget FCF	
	r_p	r_g		r_p	r_g
YLWT	-0.04 ± 0.01	-0.09 ± 0.07	HLWT	0.01 ± 0.01	-0.11 ± 0.05
YGFW	-0.09 ± 0.01	-0.21 ± 0.07	HGFW	-0.19 ± 0.01	-0.32 ± 0.04
YCFW	-0.11 ± 0.01	-0.58 ± 0.10	HCFW	-0.19 ± 0.01	-0.43 ± 0.05
YFD	-0.34 ± 0.01	-0.63 ± 0.03	HFD	-0.47 ± 0.01	-0.54 ± 0.02
YFDCV	-0.24 ± 0.01	-0.35 ± 0.05	HFDCV	-0.31 ± 0.01	-0.35 ± 0.03
YCURVE	0.08 ± 0.01	0.29 ± 0.05	HCURVE	0.21 ± 0.01	0.34 ± 0.03
YSL	-0.03 ± 0.01	-0.09 ± 0.06	HSL	-0.08 ± 0.01	-0.17 ± 0.03
YSS	0.03 ± 0.01	0.29 ± 0.12	HSS	-0.03 ± 0.02	0.06 ± 0.07
YCOL	-0.27 ± 0.10	0.73 ± 0.34			

The IN yearling and adult expressions of FCF were strongly correlated with each other, both phenotypically (0.73 ± 0.01) and genetically (0.98 ± 0.02) while the SG yearling and hogget expressions were equally highly correlated (phenotypic: 0.54 ± 0.01 and genetic: 0.97 ± 0.02). Most of the phenotypic correlations (r_p) between FCF and other key production traits in both data sets were negligible (i.e. <0.2). The exceptions were r_p with FD, FDSD, FDCV, RTOC, HAND, CHAR and SSTRC in the IN which varied in magnitude from low to high, depending on the age of expression (Table 2a) and r_p with FD, FDCV, CURVE and COL in the SG data which were all low except for HFD and HFCF which was medium (Table 2b). All of the r_p were negative indicating that animals with high FCF had finer less variable FD, lower CURVE and RTOC, increased

textural softness, well defined crimp, fine staple bundles and whiter greasy colour.

In the IN, the majority of the genetic correlations (r_g) between FCF and key production traits were significant and negative (Table 2a). The r_g between YFCF and YFD, YFDSD, AFD and AFDSD (not presented) ranged between -0.92 to -0.67 and, although those between AFCF and YFD and YFDSD were slightly lower (-0.58 and -0.57 respectively), indicate that selection to reduce FD or FDSD will generate favourable correlated increases in FCF. Interestingly the r_g between FCF and FDCV at each age, while still significant, were at least 50% lower than those involving FD or FDSD. The r_g in the SG data were reasonably consistent with the IN estimates (Table 2b) with the exception of positive r_g with YCURVE (0.29) and HCURVE (0.34), the low positive r_g between YFCF and YSS (0.29) and the high positive r_g between YFCF and YCOL (0.73). The remaining significant r_g between FCF and the various production traits in both the IN and SG data were as expected given the strong genetic correlation between FCF and FD.

The strong r_g between FCF with FD and FDCV indicate limited benefit from including FCF as an additional trait in Merino breeding programs with an existing emphasis on reducing FD, as the percentage of fibres over 30 µm is simply a reflection of FD and FDSD. This study also indicates that it would be difficult for medium to strong wool breeders to add FCF to their breeding programs without making correlated changes in FD and to a lesser extent LWT and wool production due to the underlying biology of fleece production Moore *et al*. (1996). Naebe *et al.* (2015) recently investigated the prickle response of 48 fabrics with a range of FD, yarn and fabric construction and found that including FD in the model along with other significant fibre, yarn and fabric factors rendered measures of FD distribution, including FDCV and FCF, insignificant. Therefore variation in FD together with yarn and fabric construction factors appears to have a greater impact on prickle than FCF.

ACKNOWLEDGEMENTS

We acknowledge the dedicated team of scientists and technical staff at each of the 8 IN sites who are responsible for management of each of the flocks and implementation of the comprehensive sampling, assessment and measurement protocols. This research was also funded by Meat and Livestock Australia (MLA) and Australian Wool Innovation (AWI). The authors would like to thank the ram breeders who collected data used in this study.

REFERENCES

AWI Ltd and MLA Ltd (2013) *Visual Sheep Scores - Researcher Version.* Sydney, Australia: AWI Ltd & MLA Ltd.

Baxter B.P. and Cottle D.J. (1998) *Wool Tech. Sheep Breed.* **46**: 154.

Brown D.J., Huisman A.E., Swan A.A., Graser H.U., Woolaston R.R., Ball A.J., Atkins K.D. and Banks R.G. (2007) *Assoc. Advmt Anim. Breed. Genet*. **17**: 187.

Casey A.E. and Cousins P. (2010) *Sheep Scores - Wool Handle*. Orange, Australia: NSW DPI.

Casey A.E., Martin S.J., Swain B.C. and Ramsay A.M.M. (2009) *AMSEA Accredited Central Test Sire Evaluation. Requirements for an accredited site*. Orange, Australia: NSW DPI.

Garnsworthy R.K., Gully R.L., Kandiah R.P., Kenins P., Mayfield R.J. and Westerman R.A. (1988) *Understanding the causes of prickle and itch from the skin comfort of fabrics*. Geelong, Australia: CSIRO Division of Wool Technology.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) *ASReml User Guide Release 3.0.* Hemel Hempstead, UK: VSN International Ltd.

Moore G.P.M, Jackson N., Isaacs K. and Brown G. (1996) *Aust. J. Agric. Res*. **47**:1195.

Naebe M., McGregor B.A., Swan P.G. and Tester D. (2015) *Text. Res. J.* **85**:1122.

Naylor G.R.S. (1992) *Wool Tech. Sheep Breed.* **40:** 14.

van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Anim. Prod. Sci.* **50:** 998.

Proc. Assoc. Advmt. Breed. Genet. 21: 209-212

GENETIC PARAMETERS FOR EATING QUALITY TRAITS OF AUSTRALIAN LAMB

S.I. Mortimer[1,2], A.A. Swan[1,3], L. Pannier[1,4], A.J. Ball[1,5], R.H. Jacob[1,6], J.H.J. van der Werf[1,7] and D.W. Pethick[1,4]

[1] Cooperative Research Centre for Sheep Industry Innovation, NSW, Australia
[2] NSW Department of Primary Industries, NSW, Australia
[3] Animal Genetics and Breeding Unit*, University of New England, NSW, Australia
[4] School of Veterinary and Life Sciences, Murdoch University, WA, Australia
[5] Meat and Livestock Australia, University of New England, NSW, Australia
[6] Department of Agriculture and Food, WA, Australia
[7] School of Environmental and Rural Science, University of New England, NSW, Australia

SUMMARY

Genetic parameters were estimated for 5 sensory (overall liking, tenderness, juiciness, flavour and liking of odour) and 2 objective eating quality (EQ) traits (intramuscular fat, IMF, and shear force, SF), measured on loin and topside meat cuts (except IMF) produced by progeny from the Information Nucleus of the CRC for Sheep Industry Innovation. Heritabilities for sensory traits were low to moderate for loin and moderate to high for topside cuts. Loin IMF was highly heritable while SF was moderately heritable in both cuts. Genetic correlations among the sensory EQ traits were all positive and high (0.72-1.00). Genetic correlations between loin IMF and sensory traits were moderately positive and lowly positive for loin and topside respectively. SF had stronger, negative correlations with sensory EQ traits in the topside than in the loin. Overall liking may be improved more so through selection on IMF in the loin and SF in the topside.

INTRODUCTION

Whilst breeding programs implemented in the Australian sheep industry have yielded substantial gains in lean meat yield, several studies have indicated that such programs may have also reduced the sensory eating quality (EQ) of lamb meat produced by terminal sire production systems. Breeding programs applied by breeders of terminal sires routinely use Australian Sheep Breeding Values for post weaning expressions of ultrasound eye muscle and fat depths at the C-site as components of breeding objectives and selection indices. Selection using these breeding values has been shown to reduce intramuscular fat (IMF) (e.g. Hopkins *et al.* 2007; Pannier *et al.* 2014b) and sensory EQ scores (e.g. Hopkins *et al.* 2005; Pannier *et al.* 2014a). Research is underway through the CRC for Sheep Industry Innovation to enable EQ traits to be incorporated directly into breeding programs. Contributing to this research, this study aimed to estimate genetic parameters for a range of EQ traits of samples, assessed by untrained consumer panels, from the *m. longissimus lumborum* (short loin) and *m. semimembranosus* (topside) muscles of lambs. Genetic and phenotypic correlations of the sensory EQ traits with IMF and shear force were also estimated.

MATERIALS AND METHODS

The Information Nucleus (IN) program of the CRC for Sheep Industry Innovation (van der Werf *et al.* 2010) produced the lambs that were slaughtered to obtain the meat samples for this study. The sample preparation, cooking procedures and sensory testing protocols applied to grilled loin and topside samples tasted by panels of untrained consumers have been described by Pannier *et al.* (2014a). Briefly, 10 sub-samples prepared from each meat sample of a carcass were grilled under standardised conditions during each consumer tasting session (94 sessions in total) and

*AGBU is a joint venture of NSW Department of Primary Industries and University of New England

tasted to provide 10 consumer responses per meat cut. The EQ traits were assessed by the consumers using a 0-100 scale (100 being most preferred) and for this study included tenderness (Ltend, Ttend), juiciness (Ljuic, Tjuic), liking of flavour (Lflav, Tflav), overall liking (Llike, Tlike) and liking of odour (odour) of loin and topside samples respectively. EQ records, based on the mean of 10 consumer responses for each muscle, were derived from 1414 animals, which were the progeny of 174 sires of various breeds born during 2009-2010, at 2 IN sites (Kirby and Katanning). From their carcasses and those from other animals born during 2007-2011 at all sites and slaughtered by the IN program, samples from the other loin were taken for analysis of IMF (%, 8917 records) and testing of shear force (SF) after 5 days of ageing (LSF, N, 9000 records), as described by Mortimer *et al.* (2014). A topside sample was similarly tested for SF after 5 days of ageing (TSF, N, 4793 records).

For the sensory EQ traits, univariate anlyses were used to estimate phenotypic variances and heritabilities, where models included fixed effects of site, slaughter group, sex, birth type, rearing type, age of dam and age at measurement. Random effects included a direct genetic effect of animal and an effect of consumer tasting session, as well as a fixed effect of genetic group. Univariate analyses of IMF, LSF and TSF fitted fixed and random effects in models as described by Mortimer *et al.* (2014). Genetic and phenotypic correlations among the EQ traits, IMF and shear force values were estimated from bivariate analyses. Variance and covariance estimations were performed using ASReml (Gilmour *et al.* 2014). Table 1 summarises the data for each trait.

RESULTS AND DISCUSSION

Apart from odour, phenotypic variances and heritabilities of sensory EQ traits were lower for loin compared to topside samples, although the standard errors were relatively large for the heritability estimates of both meat cuts (Table 1). Heritabilities for EQ traits of loin samples were low to moderate while those assessed on topside samples were moderate to high. Irrespective of meat cut, odour was of very low heritability. These results, though based on relatively small numbers of records, confirm that the genetic variation in EQ traits is sufficient to be exploited through selection. Previously, the much smaller study of Karamichou *et al.* (2007), using trained taste panel assessments of loin samples (350 records), had shown that equivalent lamb sensory traits to those assessed in the present study were of low (0.05±0.09 for overall liking) to high heritability (0.31±0.17 for juiciness). While IMF had a high heritability (0.53), SF when measured in either the loin (0.24) or topside samples (0.29) was of moderate heritability. These estimates were consistent with estimates from a subset of the IN data used in this study and earlier estimates reviewed by Mortimer *et al.* (2014).

Within each meat cut, genetic correlations among the sensory EQ traits were positive, with all estimates greater than 0.90 in size (Table 2, with results not shown for the odour traits due to low estimates of genetic variance and heritability). Phenotypic correlations were also high, though slightly lower than the genetic correlations. Johnston *et al.* (2003) reported that untrained panel assessments of sensory scores for tenderness, juiciness, flavour and overall acceptability of beef produced by temperate and tropically adapted cattle breeds were highly correlated, both genetically and phenotypically. The genetic correlations between the same EQ traits assessed in each of the loin and topside samples were between 0.87, for Ltend with Ttend, and 1.00, for Ljuic with Tjuic, and tended to be associated with larger standard errors. For the remaining combinations of sensory scores, assessments in loin and topside samples were positively and highly correlated genetically, ranging from 0.72 to 1.00. However, phenotypic correlations between assessments in loin and topside samples were much weaker, ranging from 0.28 to 0.36.

Very similar to the results of Karamichou *et al.* (2006) and Mortimer *et al.* (2014), IMF and LSF had a high, negative genetic correlation (-0.64, Table 2). However, a low negative genetic correlation (-0.21) was estimated between IMF measured in the loin and SF measured in the

Table 1. Descriptive statistics for lamb eating quality traits of loin and topside cuts (0-100 scale, 100 most preferred), intramuscular fat (IMF, %) and shear force after ageing for 5 days (SF, N) and estimates of phenotypic variance (σ^2_P) and direct heritability (h^2)

	Mean	SD	Minimum	Maximum	σ^2_P	h^2
Loin						
Tenderness	73.2	9.4	32.4	92.7	64.9 (2.74)	0.20 (0.07)
Juiciness	67.3	9.4	30.3	94.8	64.0 (2.67)	0.18 (0.07)
Flavour	71.1	8.0	40.5	91.2	49.5 (2.04)	0.10 (0.06)
Overall liking	72.3	8.4	40.6	93.9	53.8 (2.23)	0.15 (0.07)
Odour	69.9	6.4	46.1	87.0	32.0 (1.31)	0.04 (0.06)
IMF	4.3	1.018	1.5	10.5	0.67 (0.01)	0.53 (0.04)
SF5	27.5	9.8	10.8	95.1	58.53 (1.08)	0.24 (0.03)
Topside						
Tenderness	47.7	12.0	12.9	83.9	110.9 (4.85)	0.36 (0.09)
Juiciness	48.1	10.3	20.9	78.9	81.4 (3.45)	0.24 (0.08)
Flavour	55.2	9.4	27.8	84.4	69.9 (2.92)	0.17 (0.07)
Overall liking	52.1	10.4	20.8	84.2	85.5 (3.68)	0.30 (0.09)
Odour	66.2	6.8	44.6	84.5	36.4 (1.48)	0.03 (0.06)
SF5	41.8	12.3	15.3	98.9	84.4 (1.97)	0.29 (0.04)

Table 2. Genetic (below diagonal) and phenotypic (above diagonal) correlations among lamb eating quality traits of loin and topside cuts, intramuscular fat and shear force

	Ltend	Ljuic	Lflav	Llike	Ttend	Tjuic	Tflav	Tlike	IMF	LSF	TSF
Ltend		0.78	0.76	0.84	0.35	0.33	0.31	0.35	0.20	-0.34	-0.21
Ljuic	0.93 (0.07)		0.78	0.83	0.29	0.36	ne	0.33	0.23	-0.23	-0.17
Lflav	0.94 (0.11)	0.99 (0.10)		ne	0.28	0.31	0.34	0.34	0.22	-0.24	-0.16
Llike	0.95 (0.06)	0.90 (0.08)	ne[1]		0.32	0.34	ne	0.36	0.24	-0.28	-0.18
Ttend	0.87 (0.14)	0.72 (0.19)	0.90 (0.25)	0.93 (0.17)		0.82	ne	0.89	0.11	-0.19	-0.45
Tjuic	0.93 (0.17)	1.00 (0.18)	0.98 (0.26)	0.99 (0.19)	0.94 (0.05)		ne	0.86	0.15	-0.13	-0.34
Tflav	1.00 (0.20)	ne	0.94 (0.29)	ne	ne	ne		0.91	0.12	-0.12	-0.31
Tlike	0.88 (0.15)	0.80 (0.19)	0.90 (0.25)	0.93 (0.18)	0.98 (0.02)	0.94 (0.04)	0.99 (0.03)		0.13	-0.17	-0.41
IMF	0.48 (0.15)	0.49 (0.15)	0.60 (0.20)	0.59 (0.16)	0.15 (0.14)	0.34 (0.16)	0.36 (0.19)	0.27 (0.15)		-0.29	-0.12
LSF	-0.45 (0.17)	-0.43 (0.19)	-0.49 (0.23)	-0.44 (0.20)	-0.58 (0.15)	-0.74 (0.18)	-0.74 (0.22)	-0.60 (0.16)	-0.64 (0.06)		0.36
TSF	-0.55 (0.17)	-0.38 (0.19	-0.48 (0.23)	-0.47 (0.20)	-0.85 (0.09)	-0.80 (0.13)	-0.91 (0.15)	-0.84 (0.11)	-0.21 (0.09)	0.59 (0.08)	

[1]ne, not estimable.

topside. At the same time, the genetic correlation between SF in the loin and topside was 0.59. These results suggest that tenderness of the two muscles may be influenced by different genes. Johnston *et al.* (2001) had earlier concluded that very different genes could be influencing the expression of tenderness in the two muscles, after estimating a genetic correlation of 0.34 between

shear force in the loin and topside muscles from tropically adapted beef breeds. This was thought to be due to different levels of connective tissue in the 2 muscles as shear force measurement is an indicator of tenderness due to the myofibrillar component rather than connective tissue. For lamb cuts, total fat content of topside cuts has been reported to be lower than loin cuts (Ponnampalam *et al.* 2010) and may also be an influence.

Genetic correlations of loin IMF with EQ traits were moderately positive for the loin samples (range of 0.48 to 0.60), but lowly positive for topside samples (range of 0.15 to 0.36; Table 2), which agrees with the strong, positive genetic relationships of IMF with juiciness and flavour assessed in lamb loins reported by Karamichou *et al.* (2006). In contrast, SF was negatively correlated with EQ traits, irrespective of meat cut. Genetic correlations of LSF and TSF with sensory scores assessed on loin samples were moderate (ranges of -0.43 to -0.49 and -0.38 to -0.55), while the corresponding genetic correlations with EQ traits assessed on topside samples were much stronger (ranges of -0.58 to -0.74 and -0.80 to -0.91). Using the standard equation to estimate correlated response and assuming constant selection intensity, selection for increased IMF is predicted to yield about 100% and 70% more response in Llike than selection for reduced LSF and TSF respectively. In contrast, about 50% and 130% greater correlated responses in Tlike are likely from selection for reduced LSF and TSF respectively versus selection for increased IMF.

Conclusions. There is scope to improve sensory EQ properties of lamb loin and topside cuts through selection. Selection on EQ traits assessed on either meat cut can be expected to improve sensory traits of the other meat cut. Though using both traits in breeding programs would be beneficial, it seems that IMF may be the better indicator trait to improve overall liking of the loin, while SF, particularly assessed on the topside, could be a better indicator trait to improve consumer overall liking of the topside. The estimates of this study are preliminary and many more records are needed to obtain accurate genetic parameter estimates for EQ traits, through both pedigree- and genomics-based approaches.

ACKNOWLEDGMENTS

The CRC for Sheep Industry Innovation and all of its partners are acknowledged for their funding and support of this study.

REFERENCES

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. and Thompson R. (2014) 'ASReml User Guide Relaese 4.0'. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Hopkins D.L., Hegarty R.S. and Farrell T.C. (2005) *Aust. J. Exp. Agric.* **45**: 525.

Hopkins D.L., Stanley D.F., Toohey E.S., Gardner G.E., Pethick D.W., *et al.* (2007) *Aust. J. Exp. Agric.* **47**: 1219.

Johnston D.J., Reverter A., Robinson D.L. and Ferguson D.M. (2001) *Aust. J. Exp. Agric.* **41**: 991.

Johnston D.J., Reverter A., Ferguson D.M., Thompson J.M. and Burrow H.M. (2003) *Aust. J. Agric. Res.* **54**: 135.

Karamichou E., Richardson R.I., Nute G.R., McLean K.A. and Bishop S.C. (2006) *Anim. Sci.* **82**: 151.

Karamichou E., Richardson R.I., Nute G.R., Wood J.D. and Bishop S.C. (2007) *Animal* **1**: 1524.

Mortimer S.I., van der Werf, J.H.J., Jacob, R.H., Hopkins, D.L., Pannier L., *et al.* (2014) *Meat Sci.* **96**: 1016.

Pannier L., Gardner G.E., Pearce K.L., McDonagh M., Ball A.J., *et al.* (2014a) *Meat Sci.* **96**: 1076.

Pannier L., Pethick D.W., Geesink G.H., Ball A.J., Jacob R.H., *et al.* (2014b) *Meat Sci.* **96**: 1068.

Ponnampalam E.N., Warner R.D., Kitessa S., McDonagh M.B., Pethick D.W., *et al.* (2010) *Anim. Prod. Sci.* **50**: 775.

van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Anim. Prod. Sci.* **50**: 998.

Proc. Assoc. Advmt. Breed. Genet. 21: 213-216

A PUBLIC:PRIVATE COLLABORATION TO EVALUATE THE POTENTIAL VALUE OF GENOMIC INFORMATION TO A VERTICALLY-INTEGRATED COMMERCIAL BEEF CATTLE ENTERPRISE

B.T Welly[1], M.D. MacNeil[2], A.R. McClain[3], R.C. Raymond[3], E.B. Simpson[4], and A.L. Van Eenennaam [1]

[1] Department of Animal Science, University of California, Davis, CA, USA
[2] Delta G, Miles City, MT, USA
[3] J. R. Simplot Land and Livestock, Grand View, ID, USA
[4] GeneSeek, a Neogen company, Lincoln, NE, USA

SUMMARY

The objective of this collaborative research project was to use phenotypes collected from Charolais-sired crossbred calves in a commercial feedlot and processing plant to develop genomically-enhanced EBVs. Phenotypes and genotypes were collected from 4,195 crossbred feedlot calves and genomic breeding values (GBV) were calculated for post-weaning average daily gain, hot-carcass weight, marbling (MRB), ribeye area, and external fat thickness (FAT). Estimated breeding values (EBV) for Charolais sires with 10 or more progeny were calculated using an animal model with MTDFREML. Correlations of GBV with EBV ranged from 0.84 to 0.93 when all calves were included in the data, but dropped to between 0.13 and 0.31 when sire's own progeny were removed from the data set using a 5-fold cross-validation approach. Correlations increased when narrowing the evaluation to only those sires with 15 or more progeny, resulting in trait GBV accuracies ranging from 0.18 to 0.45 for FAT and MRB, respectively. The inclusion of additional progeny in subsequent years of this project is expected to improve the accuracies of genomic predictions, and data will be used to evaluate the potential uses, costs and predicted benefits of using genomic information to optimize breeding program design and management on this vertically-integrated beef operation.

INTRODUCTION

To explore the potential economic value of genomic information to a large, vertically-integrated beef cattle enterprise, a collaborative research agreement was formed between J.R. Simplot Land and Livestock and their genetic consultant Dr. Michael MacNeil, Delta G; the University of California, Davis; and Neogen/GeneSeek. Objectives of the project are i) to develop genomically-enhanced EBVs using data collected from commercial calves in the feedlot for the selection of terminal sire seedstock, ii) determine the cost:benefit of incorporating genomics into seedstock selection for an enterprise that derives value improvement in feedlot and processor economically-relevant traits, and iii) examine other opportunities for deriving additional value from the genomic information such as marker-assisted management of the feedlot calves or an optimized breeding program design for this enterprise. Here we provide a preliminary report of results from the first year of data collection.

MATERIALS AND METHODS

Phenotypes were collected from 4195 crossbred feedlot calves sired by Charolais terminal sires. Crossbred calves were genotyped with the "Low Density GeneSeek Genomic Profiler" (**GGP_LD)** bead chip that includes 19,725 SNPs. The phenotypic data collected for this project includes sex and polled status and encompasses traits involved in feedlot performance and carcass merit. Current traits analyzed for this project include post-weaning average daily gain (**ADG**), hot-carcass weight (**CWT**), marbling (**MRB**), ribeye area (**REA**), and backfat thickness (**FAT**). A

total of 629 Charolais sires were genotyped, 415 with the "High Density GeneSeek Genomic Profiler" (**GGP_HD**) that includes 76,883 SNPs and 214 with the GGP_LD. Sire assignment was performed on calves with all sires as potential candidates for each run using an exclusion-based method implemented by SireMatch (J. Pollak, Cornell University). Two sets of 500 SNPs, selected based on high minor allele frequency (**MAF**) and high call rate, were utilized for the assignment of sires.

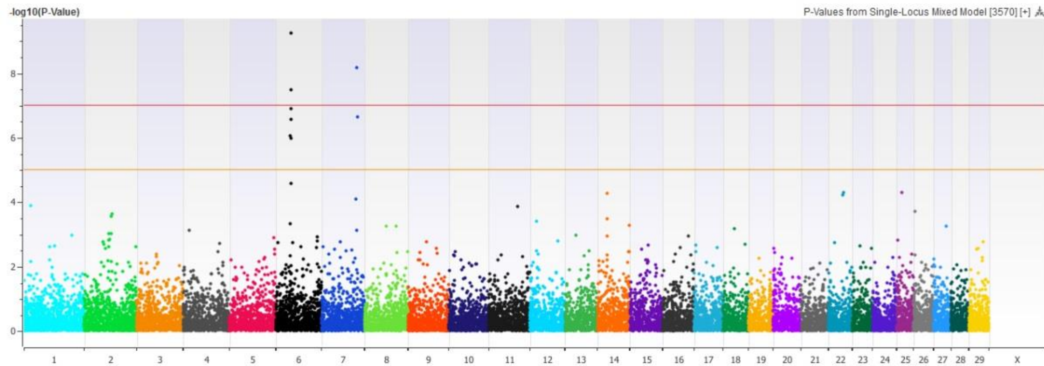
A genome wide association study was performed on all traits using the Efficient Mixed-Model Association eXpedited (EMMAX) model implemented in Golden Helix. The mixed model equations consisted of contemporary groups based on sex, ranch origin (10 ranches), and harvest date. The GWAS was carried out utilizing phenotypes on 3,555 crossbred calves and the 15,658 SNP from the GGP_LD. SNPs surrounding significant QTL peaks were extracted to evaluate the proportion of genetic variance explained by SNPs in the QTL region.

Estimated breeding values (**EBV**) and heritability ($\mathbf{h^2}$) estimates were first calculated using an animal model with MTDFREML (Boldman *et al.*, 1995). Due to a lack of pedigree data, sires (established via genotyping) were considered unrelated and dams were unknown. The EBV for Charolais sires with 10 or more progeny were extracted from the results. The GBLUP method implemented in Golden Helix's SNP and Variation Suite (Golden Helix, Bozeman, MT) was then utilized to estimate SNP marker effects on 8,000 SNP that are common to both the GGP_LD and GG_HD, for prediction of genomic breeding values (**GBV**). A 5-fold cross validation approach was used to calculate the accuracy of the GBV. Sires with EBV were randomly allocated to one of 5 groups such that approximately an equal number of progeny were included in each group. In each training analysis, the progeny of the sires in each of the 5-fold cross validation groups were excluded for the development of the genomic prediction equation for those sires. Accuracy of the genomic breeding value for the sires was calculated as the correlation between the EBV and the GBV, divided by the average accuracy of the EBV (Neves *et al.* 2014).

RESULTS AND DISCUSSION

Collection of phenotypes at the feedlot and through the processing facility was facilitated by the use of electronic capture of all records at the processing chute, and the use of matched pair sets of visual ID and EID and the nextGen™ ear tissue sampling unit (Allflex USA, Dallas, TX) to collect a DNA sample for genotyping. A total of 4195 DNA samples were analysed of which 3269 were identified to a total of 325 single sires (77.93%). The use of two sets of 500 SNPs for sire exclusion clearly identified animals with no genotyped sire. Principal component analysis of the genomic data clearly revealed clustering of half-sib groups for groups of calves with no sire assignment. Data from calves that were not assigned to a specific sire were also used as part of the training population for the GBV. Collection of DNA from all possible sires remains one of the predominant difficulties when working with large commercial populations. The proportion of possible sires that were genotyped increased for the year 2 progeny as demonstrated by an increase in sire assignment rate to 87.5%.

The GWAS analysis identified a small number of calves with incorrect gender assignment, and correctly identified a significant LOD peak for the polled locus on Chromosome 1. Significant SNPs were identified for CWT, ADG, and FAT on chromosomes 6 and 7 (**Figure 1**), both of which have been associated with growth in beef cattle previously (Lindholm-Perry *et al.*, 2011; Saatchi *et al.*, 2014). Further analysis for CWT on SNPs surrounding the peak on chromosome 6 revealed that 10 SNP on either side of the peak accounted for 6.29% of the genetic variance with the most significant SNP accounting for 1.1%. Two SNP in close proximity to the peak on chromosome 7, and two SNP upstream accounted for 1.72% and 0.77% of the genetic variance, respectively. The identification of QTLs that are in common with those found in other studies using different breeds of cattle supports the integrity of the field phenotypic data collection in this



study.

Figure 1. Manhattan plot using 3555 hot carcass weight (CWT) records implemented with a mixed model analysis, using the Efficient Mixed-Model Association eXpedited (EMMAX) model.

Accuracies of genomic prediction using a 5-fold cross validation approach ranged from 0.13 to 0.38 and from 0.18 to 0.45 for sires with ≥10 or ≥ 15 progeny, respectively (**Table 1**). Estimates of GBV accuracy using this 5-fold cross-validation approach are likely lower than true accuracy, because a large number of calves (i.e. ≥10 calves x ~ 22 sires) were removed from the training population in each of the five iterations.

Table 1. Accuracies of EBVs of Charolais sires with at least 10 progeny records, and GBVs[1] for the same sires when using all progeny records to train the prediction equations, or when excluding the sire's own progeny from the training population using a 5-fold cross validation approach.

			Sires ≥ 10 Progeny					Sires ≥ 15 Progeny				
Trait[2]	h^2 ± SE	N_{Train}[3]	N_{Sire}[4]	EBV Acc.[5]	r_{All}[6]	$r_{5\text{-}Fold}$[7]	GBV Acc.[8]	N_{Sire}[4]	EBV Acc.[5]	r_{All}[6]	$r_{5\text{-}Fold}$[7]	GBV Acc.[8]
ADG	0.31±0.06	3392	112	0.75	0.84	0.21	0.29	65	0.78	0.88	0.29	0.37
CWT	0.32±0.06	3555	114	0.74	0.92	0.22	0.30	72	0.77	0.92	0.21	0.28
MRB	0.49±0.08	3370	111	0.80	0.92	0.31	0.38	67	0.82	0.93	0.37	0.45
REA	0.40±0.07	3370	111	0.77	0.87	0.21	0.28	67	0.8	0.89	0.27	0.33
FAT	0.49±0.08	3370	111	0.80	0.93	0.13	0.16	67	0.82	0.94	0.15	0.18

[1]Genomic breeding value (GBV) accuracy estimates were calculated on Charolais sires with at least 10 or at least 15 progeny records; [2]ADG = average daily gain from feedlot arrival to final implant (μ= 135 days); CWT= carcass weight; MRB= marbling scored by camera; REA= ribeye area scored by camera; FAT= external fat thickness in adjusted yield grade units; [3]Number of crossbred calves with associated phenotypes and genotypes used to train the prediction equations; [4]Number of Charolais sires with ≥ 10 and 15 progeny respectively; [5]Average accuracy of estimated breeding values (EBV); [6]Pearson's correlation between EBV and GBV, r(EBV,GBV), when all crossbred calves are included in the training; [7]Pearsons correlation between EBV and GBV, r(EBV,GBV), for 5-fold cross-validation, where progeny from one sire group were excluded for the prediction of GBVs for that respective sire group; [8]Accuracies calculated as the Pearson's correlation between the EBV and the 5-fold cross-validated GBV, divided by the average accuracy

of the EBV.
Accuracy of genomic prediction when using phenotypes is affected by heritability of the trait, quality of the phenotypic data, number of animals in the training population for each trait, marker density, and statistical prediction methodology. Thus, we anticipate improved accuracy to result from increases in the number of sires and phenotyped calves from subsequent calf drops, as well as future work to impute genotypes to greater density (Marchini and Howie, 2010), and implement Bayesian prediction methodology (Fernando *et al*., 2014).

The impediments to the adoption of genomic technology in the beef cattle industry include the need for large training populations, the lack of a national breeding objective that includes and appropriately weights varying economic drivers in the different sectors of the beef cattle industry, and the difficulty of obtaining phenotypes from the whole supply chain. Much of the value derived from selection at the seedstock sector is realized by downstream supply chain partners (e.g. processing sector). Frequently there is no price signalling back to the seedstock producer making investments in phenotyping and genotyping to improve genetic progress in these traits, and this market failure impacts the commercial viability of any genetic technology (Van Eenennaam *et al*., 2011).

Vertically-integrated enterprises have the opportunity to develop their own breeding objective, and derive all of the value associated with genetic improvements across the various sectors of the beef industry, and hence are ideally situated to fully realise the potential of genomic information (Van Eenennaam and Drake, 2012). One advantage that vertically-integrated beef operations have when developing their breeding objective is the opportunity to include non-conventional traits. They are more likely to have ready access to records of economically relevant traits (e.g. feedlot feed requirements; survival to market endpoint) with very high relative economic value (Van Eenennaam and MacNeil, 2011), or related indicator traits (e.g. disease treatment/death records).

It is envisioned that at the end of this three-year collaborative project accurate GBVs will have been developed for traits of economic importance to this large vertically-integrated beef cattle enterprise for their Charolais terminal sire seedstock herd, and the value proposition associated with the multiple potential uses of the genomic information and phenotypic information being collected as a part of this project will have been evaluated.

ACKNOWLEDGEMENTS

The authors acknowledge the excellent field data collection by Tommie L. Crockett, Joel A. Ontiveros, and Nicole Sykes.

REFERENCES

Boldman K.G., Kriese L.A., Van Vleck L.D., Van Tassell C.P. and Kachman S.D. 1995. A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances. U.S. Department of Agriculture, Agricultural Research Service. Washington, DC.

Fernando R.L., Dekkers J.C., and Garrick D.J. (2014) *Genet. Sel. Evol*. **46**:50.

Lindholm-Perry A.K., Sexten A.K., Kuehn L.A., Smith T.P., King D.A., *et al*. (2011) *BMC Genet*. **14**:103.

Marchini, J. and Howie, B. (2010). *Nat. Rev. Genet.* **11**:499.

Neves H.H., Carvalheiro R., O'Brien A.M., Utsunomiya Y.T., do Carmo A.S., **et al**. (2014) *Genet. Sel. Evol.* **46**:17.

Saatchi, M., Schnabel R.D., Taylor J.F., and Garrick, D.J. (2014) *BMC Genomics* **15**: 442.

Van Eenennaam A.L., and Drake D.J. (2012) *Livest. Prod. Sci.* **52**:185.

Van Eenennaam, A.L. and MacNeil M.D. (2011) pp.61-68 in "Proceedings of the Beef Improvement Federation 43rd Annual Research Symposium". Bozeman, MO.

Van Eenennaam A.L., van der Werf J.H.J. and Goddard M.E. (2011) *J. Anim. Sci.* **89**:307.

Proc. Assoc. Advmt. Breed. Genet. 21: 217-220

PRODUCTION PERFORMANCE OF MERINO AND DOHNE MERINO EWES AND LAMBS IN PURE OR CROSSBREEDING SYSTEMS

S.W.P. Cloete[1,2] and J.J.E. Cloete[1,3]

[1]Department of Animal Sciences, Stellenbosch University, Matieland 7602, South Africa
[2]Directorate Animal Sciences: Elsenburg, Private Bag X1, Elsenburg 7607, South Africa
[3]Elsenburg Agricultural Training Institute, Elsenburg 7607, South Africa

SUMMARY

This study details the performance of Merino and Dohne Merino ewes and lambs over eight years in either a pure or a crossbreeding regime. Ewes were mated either to rams of their own breed or to Dormer or Suffolk rams as terminal sires. Dam breed affected birth weight and lamb survival, with lambs borne by Dohne ewes being heavier and having greater survival to weaning than those borne by Merinos. Progeny of Dohne ewes and progeny sired by terminal rams were heavier at weaning. Dohne Merino ewes were heavier at mating than Merinos, but produced less clean wool at a slightly lower fibre diameter. No ewe breed or breeding regime differences were found for number of lambs born or weaned per ewe lambed. Total weight of lamb weaned was higher in Dohne Merino ewes and in ewes mated to terminal rams. Crossbreeding may have a relative advantage to pure breeding in terms of lamb output per unit ewe body weight maintained. Further studies on breed differences and crossbreeding of the South African ovine genetic resource are warranted.

INTRODUCTION

Research on South African sheep has so far not focused on the comparison of those breeds constituting the available ovine genetic resource. The option of crossbreeding as a means to achieve commercial gains through heterosis and the exploitation of sexual dimorphism has also not been researched in great detail. In fact, published studies on these topics are very scarce.

The paucity of published research is not only relevant for the South African sheep industry, as South African ovine germplasm has been exported to several overseas countries, including Australia and New Zealand. Among the breeds that were exported to Australia is the Dohne Merino. This breed presently contributes the most weaning weight records to the South African national database and shows sustained growth in weaning weight records during the recent decade (Cloete *et al*. 2014). In Australia, the Sheep Genetics Database already includes more than 100,000 Dohne Merino records for most key traits and the breed is regarded as adaptable, with easy-care properties and an ability to adapt to varying conditions (Li *et al*. 2013).

Against this background, we assessed the performance of the Dohne Merino in comparison with the internationally known Merino breed. The breeds were compared under regimes involving pure breeding and crossbreeding with terminal sires under commercial conditions.

MATERIALS AND METHODS

The study took place from 2007 to 2014 on the Langgewens research farm near Malmesbury in the Swartland region of South Africa, a mixed farming region (grain-growing and sheep farming) described in Cloete *et al*. (2003; 2004). The Merinos used in the study included some animals used in previous crossbreeding studies (Cloete *et al*. 2003; 2004) and ewes originating from the "Wet and dry" line at Tygerhoek (Cloete *et al*. 2007). Although Merino ewes originated from experimental flocks, previous results were consistent with other reports on other industry flocks (Cloete *et al*. 2003) while some ewes born from 2007 were sired by industry rams. The Dohne ewes were also transferred from the previous experiment, but were complemented with ewes

donated by local breeders and purchased ewes, primarily from the University of Stellenbosch Dohne Merino stud (Cloete *et al*. 1999). The ewes were mated in single sire groups in January, either to rams of their own breed (n=7 for Dohne Merinos; n=12 for Merinos), or at random to rams of the terminal sire breeds Dormer or Suffolk (Cloete *et al*. 2003; 2004). The ewes were maintained as a single flock afterwards. Selection of most rams considered their representation as sires in several industry flocks so as to create links of the experimental population with the national database. As well, about half of the selected Merino rams originated from the High line of a divergent selection experiment for number of lambs weaned per mating (Cloete *et al*. 2009). The ewes lambed in June-July. Birth weight, dam identity and pedigree were recorded at birth as reported by Cloete *et al*. (2003). Weaning weights were recorded at 97 (s.d.=19) days and birth and weaning records were combined to derive complete reproduction records. Ewes were shorn in May during late pregnancy and individual greasy fleece weights were recorded. Individual wool samples were taken to measure clean yield, staple length, staple strength, fibre diameter and the coefficient of variation (CV) of fibre diameter. Clean fleece weight was derived from the product of greasy fleece weight and clean yield. Wool records were available from 2008 to 2014.

Data were analysed by ASREML (Gilmour *et al*. 2006) to predict means for selected fixed effects. Fixed effects assessed for lamb records were ewe breed (Dohne Merino or Merino), breeding regime (pure or terminal cross), sex (male or female), birth year (2007-2014), dam age (2-7+years) and birth type (single or multiple). Apart from ewe breed and breeding regime, ewe age (2-7+years) and year (2007-2014) were fitted to ewe records. No distinction was made between the two terminal sire breeds, as they were earlier shown to perform alike (Cloete *et al*. 2003; 2004). The ewe breed x breeding regime interaction was fitted where appropriate (i.e. lamb traits and ewe reproduction) but not for ewe wool traits. The random effects of animal and dam permanent environment were fitted to lamb records, while animal permanent environment (and service sire for reproduction records) were fitted to ewe traits. Where proportions needed to be analysed, the online tool of Preacher (2001) was used.

RESULTS AND DISCUSSION

Ewe breed exerted a marked effect on lamb birth weight, lambs borne by Dohne Merino ewes being 9.6% heavier (P<0.01) than the progeny of Merino ewes (5.36±0.09 vs. 4.89±0.08 kg respectively; Table 1), when expressed relative to Merino progeny. Birth weight was independent of crossbreeding system and the interaction between ewe breed and crossbreeding regime. In a previous study, when only terminal crossbreeding was practiced, progeny of Dohne Merino ewes were 5.5% heavier than lambs borne by Merino ewes (Cloete *et al*. 2003).

Table 1. Predicted means (±s.e.) for ewe breed (Merino or Dohne Merino) with mating system (pure breeding of terminal cross) combinations for lamb birth weight, weaning weight and lamb survival. The logit transformation was applied to binomial survival records, but only backtransformed means and approximate s.e.'s are given.

Ewe breed (EB)	Merino		Dohne Merino		EB	CS	EB x CS
Crossing system (CS)	Pure	Cross	Pure	Cross			
Lambs born (n)	608	395	366	312			
Birth weight (kg)	4.79±0.11	4.99±0.09	5.32±0.13	5.40±0.10	**	0.13	0.41
Weaning weight (kg)	29.1±0.5	31.9±0.5	34.4±0.6	35.5±0.5	**	**	*
Lamb survival	0.78±0.02	0.76±0.03	0.81±0.03	0.83±0.03	*	0.81	0.40

* P<0.05; ** P<0.01; Actual significance level for P>0.05

Weaning weight was affected by ewe breed, crossbreeding regime and their interaction. Terminal crossbred progeny of Merino ewes were 9.6% heavier at weaning than purebred lambs

(P<0.05; Table 1). This difference was smaller for Dohne Merino ewes, the advantage of terminal crossbred progeny only amounting to 3.2%. Overall, crossbred progeny of Dohne Merino ewes were 11.2% heavier than crossbred lambs produced by Merino ewes (Table 1). Cloete *et al*. (2003) previously reported a comparable breed difference of 10.5%. Van Deem *et al*. (2008) also found that F1 Dohne Merino x Merino lambs outgrew purebred Australian Merino lambs. Only ewe breed affected lamb survival, being improved by 6.5% in the progeny of Dohne Merinos compared to Merinos (0.82±0.02 vs. 0.77±0.02 respectively), when expressed relative to the mean for lambs borne by Merinos. A previous study reported respective lamb mortality rates of 0.18 to 0.23 for Merinos, compared to 0.16 for Dohne Merinos (Cloete *et al*. 2003). Cloete *et al*. (1999) reported that the advantage in lamb survival of purebred Dohne Merino lambs relative to Merinos was primarily for the survival of twins (respectively 0.87 vs. 0.81).

Ewe mating weight and reproduction were independent of crossing regime and the interaction of ewe breed with crossing regime (Table 2). Overall, Dohne Merino ewes were 19.6% heavier than Merinos at mating (72.6±0.6 vs. 60.7±0.4 kg respectively), when expressed relative to means for the Merino. Previous studies also reported that mature Dohne Merino ewes were approximately 20% heavier than Merinos (Cloete *et al*. 2003; 2004). The number of lambs born per ewe lambed ranged from 1.56 to 1.63 and number of lambs weaned per ewe lambed from 1.19 to 1.29 (both P>0.05). Overall, lamb output per ewe lambed of Dohne ewes exceeded the mean performance of Merinos by 18.5% (48.0±1.2 vs. 40.5±1.0 kg respectively; Table 2). The corresponding advantage of ewes mated to terminal sire rams amounted to 15.6% (47.5±1.2 vs. 41.1±1.0 kg respectively). Reproduction was not expressed per ewe mated, as higher lambing rates were seen in ewes mated to a terminal sire. In total, 395 of 512 Merino ewes mated to a Merino ram lambed in comparison to 249 of 295 ewes mated to a terminal ram (0.771 vs. 0.844; Chi^2=6.12; P=0.013). A total of 237 of 320 Dohne ewes mated to a Dohne ram lambed in comparison with 204 of 232 Dohne ewes mated to a terminal ram (0.741 vs. 0.879; Chi^2=16.1; P<0.01). Previous studies also did not report marked advantages in reproduction traits of either breed in comparison to the other in either pure or crossbred situations (Basson *et al*. 1969; Fourie and Cloete 1993; Cloete *et al*. 2003).

Table 2. Predicted means (±s.e.) for ewe breed (Merino or Dohne Merino) with mating regime (pure breeding of terminal cross) combinations for ewe mating weight and reproduction, with all reproduction traits expressed per ewe lambed

Ewe breed (EB)	Merino		Dohne		EB	CS	EB x CS
Crossing system (CS)	Pure	Cross	Pure	Cross			
Ewes lambed (n)	395	249	237	204			
Mating weight (kg)	60.7±0.5	60.8±0.5	72.5±0.6	72.7±0.6	**	0.83	0.99
Lambs born	1.58±0.03	1.63±0.04	1.56±0.04	1.59±0.04	0.42	0.19	0.51
Lambs weaned	1.19±0.04	1.20±0.05	1.23±0.05	1.29±0.05	0.20	0.46	0.56
Weight weaned (kg)	37.4±1.2	43.5±1.5	44.7±1.5	51.4±1.7	**	**	0.85

* P<0.05; ** P<0.01; Actual significance level for P>0.05

Wool traits of adult Merino and Dohne Merino ewes are provided in Table 3. Clean fleece weights of Dohne Merino ewes were 9.7% lower than in Merinos while the clean yield of Dohne Merino ewes were 7.2% below that of Merinos, when expressed relative to means for the Merino (all P<0.01). Previous studies suggested somewhat higher advantages in clean fleece weight for Merino ewes relative to Dohnes, ranging from 18 to 29% (Cloete *et al*. 1999; 2003; 2004). Van Deem *et al*. (2008) similarly reported that clean wool production was improved in Merino lambs compared to to F1 Dohne x Merino lambs. Clean yield results confirmed previous results that the clean yield of Merino ewes are higher compared to Dohnes (Basson *et al*. 1969; Cloete *et al*. 1999; 2003; Van Beem *et al*. 2008). Staple length was independent of ewe breed, but staple strength

tended to be higher in Merinos (P=0.055). Cloete *et al*. (2003) reported a mean staple strength of 37.2 N/ktex for Dohne Merino ewes compared to 41.6 N/ktex for Merino ewes selected for fleece weight. Fibre diameter and the CV of fibre diameter were improved by respectively 3.2 and 7.1% in Dohne ewes (P<0.05). Previous studies on fibre diameter are inconclusive for comparison of Dohnes with Merinos. Cloete *et al*. (2003; 2004) reported that Merino wool was broader than Dohne wool, Cloete *et al*. (1999) reported no breed difference and van Beem *et al*. (2008) reported that F1 Dohne crossbred lambs produced broader wool than Merinos.

Table 3. Predicted means (±s.e.) for ewe breed (Merino or Dohne Merino) for ewe wool traits

Trait	Breed		Significance
	Merino	Dohne	
Ewes shorn (n)	472	291	
Clean fleece weight (kg)	3.91±0.04	3.53±0.05	**
Clean yield (%)	69.9±0.2	64.9±0.3	**
Staple length (mm)	89.9±0.5	90.1±0.6	0.47
Staple strength (N/ktex)	36.2±0.5	35.1±0.6	0.06
Fibre diameter (µm)	21.6±0.1	20.9±0.1	**
Coefficient of variation (%)	19.7±0.1	18.3±0.2	**

* P<0.05; ** P<0.01; Actual significance level for P>0.05

CONCLUSIONS

There were clear advantages of Dohne Merinos for growth, lamb survival and mature size, while Merinos outperformed Dohne Merinos for clean fleece weight and clean yield. No conclusive advantage for either breed was seen in reproduction. The finer fibre diameter of Dohne Merino ewes was unexpected, but consistent with some literature. The experimental outlay did not allow the unbiased estimation of heterosis but the larger crossbred advantage for weaning weight of Merino progeny compared to Dohnes may originate from heterosis. Alternatively, the improved outputs of the crossbreeding regime could simply stem from differences in sexual dimorphism between sire and dam breeds. Further research on input/output performances are warranted.

REFERENCES

Basson W.D., Van Niekerk B.D.H., Mulder A.M. and Cloete J.G. (1969) *Proc. S. Afr. Soc. Anim. Prod.* **8**: 149.

Cloete S.W.P., Cloete J.J.E., Durand A. and Hoffman L.C. (2003) *S. Afr. J. Anim. Sci.* **23**: 223.

Cloete S.W.P., Cloete J.J.E., Herselman M.J. and Hoffman L.C. (2004) *S. Afr. J. Anim. Sci.* **34**: 135.

Cloete S.W.P., Coetzee J., Schoeman S.J., Morris J. and Ten Hoope J.M. (1999) *Proc. Assoc. Advmt. Anim. Breed. Gen.* **13**: 181.

Cloete S.W.P., Misztal I. and Olivier J.J. (2009) *J. Anim. Sci.* **87**: 2196.

Cloete S.W.P., Olivier J.J., du Toit E. and Dreyer F.H. (2007) *S. Afr. J. Anim. Sci.* **37**: 237.

Cloete, S.W.P., Olivier, J.J., Sandenbergh, L. and Snyman, M.A. (2014) *S. Afr. J. Anim. Sci.* **44**: 308.

Fourie A.J. and Cloete S.W.P. (1993) *S. Afr. J. Anim. Sci.* **23**: 104.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2006) ASREML - User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP11ES, UK.

Li L., Brown D.J. and Gill J.S. (2013) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **20**: 241.

Preacher K.J. (2001) Available from http://quantpsy.org.

Van Beem D., Wellington D., Paganoni B.L., Vercoe P.E. and Milton, J.T.B. (2008) *Aust. J. Exp. Agric.* **48**: 879.

Proc. Assoc. Advmt. Breed. Genet. 21: 221-224

HAS THE BEEF GENETIC IMPROVEMENT PIPELINE BEEN EFFECTIVE?

P.F Parnell

Angus Australia, 86 Glen Innes Road, Armidale, NSW, Australia

SUMMARY

Genetic improvement achieved by the Angus breed was examined to demonstrate the effectiveness of the beef genetic delivery pipeline in Australia. This pipeline has resulted in superior rates of genetic improvement for key economic traits relative to those achieved for the Angus breed in other countries. The accumulated present value of returns in the temperate Australian beef industry resulting from the genetic improvement achieved in Angus to 2014 was estimated to be $1.621 billion, projected to increase to $2.514 billion in 2024.

INTRODUCTION

There has been debate in recent times regarding the effectiveness of the existing pipeline for the delivery of genetic improvement to the Australia beef industry (e.g. Woolaston, 2014). The current pipeline, described by Parnell (2007), involves partnerships between cattle breeders, breed associations, commercialisers, and RD&E providers. Genetic evaluation using BREEDPLAN (Graser *et al* 2005) has been a central element of the pipeline over the past 30 years. Hammond (2006) reflected on the past success of collaborative partnership arrangements between Meat and Livestock Australia (MLA), the Animal Genetics and Breeding Unit (AGBU), the Agricultural Business Research Institute (ABRI), state departments of agriculture, various breed associations and pioneering breeders in the design, development and implementation of BREEDPLAN and its important enhancement, BreedObject (Barwick and Henzell 2005).

This paper examines genetic improvement in the Angus breed as an example of what has been achieved through the existing delivery pipeline. Members of Angus Australia are responsible for over 40% of animals registered in the Australian beef seedstock sector, as compiled by the Australian Registered Cattle Breeders Association (ARCBA). They account for 40-60% of the performance records collected in the sector. The Angus database contains over 1.75 million animals, of which 1.2 million animals have performance data contributing to a total of over 6.7 million records. Angus Australia makes considerable annual investment in breed development initiatives including provision of recording and genetic evaluation services to members, pedigree and DNA quality assurance, and the conduct of applied research. An important component of Angus Australia's business is its partnerships with MLA, AGBU, ABRI and other service providers in the provision genetic evaluation services for its members.

MATERIALS AND METHODS

In order to compare rates of genetic improvement in the Australia Angus population with those achieved in other countries the published trends in Estimated Breeding Values (EBVs) and/or Expected Progeny Differences (EPDs) for a sample of comparable traits were scaled to standard units. The Australian, New Zealand and UK EBV trends were scaled according to the respective additive genetic variances assumed in the Angus BREEDPLAN analysis. The USA and Canadian EPD trends were scaled according to the additive genetic variances used to "import" these EPDs into the Angus BREEDPLAN analysis (B.Tier, pers. comm.). These trend comparisons can only be considered approximate due to differences in trait definitions across countries and differences in the analysis models used in each respective analysis. For example, a full multi-trait model is used to calculate EBVs in BREEDPLAN, whilst some traits (e.g. birth weight) are only included in a single-trait analysis model to calculate US and Canadian EPDs.

A simple model was developed to quantify the approximate economic gains achieved in the temperate Australian beef industry resulting from genetic improvement by Angus seedstock breeders since the commencement of Angus BREEDPLAN genetic evaluation in the mid-1980s. The model accounted for genetic improvement in the registered Angus seedstock population, as measured by the trend in the average Angus Breeding Index value (Angus Australia, 2015). For simplicity, the model ignored any benefits accrued in the sub-tropical northern beef industry where Angus genetics has also had significant penetration in recent years.

Since no accurate statistics are available on the breed composition in the Australian beef herd it was assumed that the proportion of Angus animals represented in the temperate commercial beef population was equivalent to the proportion of Angus cattle in the seedstock sector relative to the total number of breeding females registered across all temperate breeds, as published annually by ARCBA. These statistics show that the proportion of Angus cattle in the seedstock sector increased from about 10% in the early 1980s to 47% in 2014. The model assumed that 75% of the industry sources their bull replacements from recorded seedstock herds (Tier 1 commercial herds), with a 5 year lag in genetic improvement (approximately 1 generation). Further, it was assumed that the remaining 25% of commercial herds (Tier 2 commercial herds) had a 10 year lag (approximately 2 generations) in genetic improvement relative to the seedstock sector.

Statistics on the numbers of breeding females mated each year in the temperate Australian beef herd were estimated from industry data provided by MLA (B. Thomas, pers. com.), with 50% of these cows assumed to be run in temperate regions where Angus bulls are commonly used. A discount rate of 7% was used to adjust returns over time to present value. Predicted returns for the subsequent 10 years beyond 2014 assumed no change in the size of the commercial cattle population and no further increase in the proportion of Angus cattle in the temperate beef herd.

RESULTS AND DISCUSSION

Figure 1 shows that the average rates of genetic improvement in key economic traits achieved in the Angus bred in Australia over the past two decades generally exceeds those achieved in other major Angus populations globally. Unfortunately, it is not possible to compare progress in profitability indexes used in different countries due to vastly different index assumptions used. However, it is expected that comparative gains in overall profitability will be directly related to gains achieved in key economic traits as shown.

Figure 2 shows the actual trends in average Angus Breeding Index values for registered Angus herds to 2014, and the predicted lagged trends in Tier 1 and Tier 2 commercial herds in temperate Australia. Also shown are the discounted returns over time from the modelled genetic trends in temperate commercial herds. These results show that the long-term genetic improvement in Angus seedstock herds, coupled with increased market share of Angus, has generated significant economic benefits for the commercial beef industry in Australia. The accumulated present value of returns resulting from the genetic improvement to 2014 was estimated to be $1.62 billion, projected to increase to $2.51 billion in 2024. Considering the market share of Angus at the time, these results are consistent with the magnitude of cumulative gross returns resulting from selection and crossbreeding in southern Australia to 2001, estimated by Farquarson *et al.* (2003).

Breeders typically have a myriad of breeding objectives, based on their interpretation of market signals and factors influencing profitability. Consequently, it is inaccurate to assume that all participants of the value chain agree with the implied breeding objectives underpinning the trends in index values typically used to monitor genetic progress in the industry. Angus breeders who have focused on profitability indexes have made excellent genetic progress with respect to these indexes (Johnston 2007). Sub-potential gains achieved in some herds or breeds are likely not to be a function of the delivery pipeline, but rather the lack of motivation for genetic change by some breeders and/or differing breeding objectives not adequately described by existing measures.

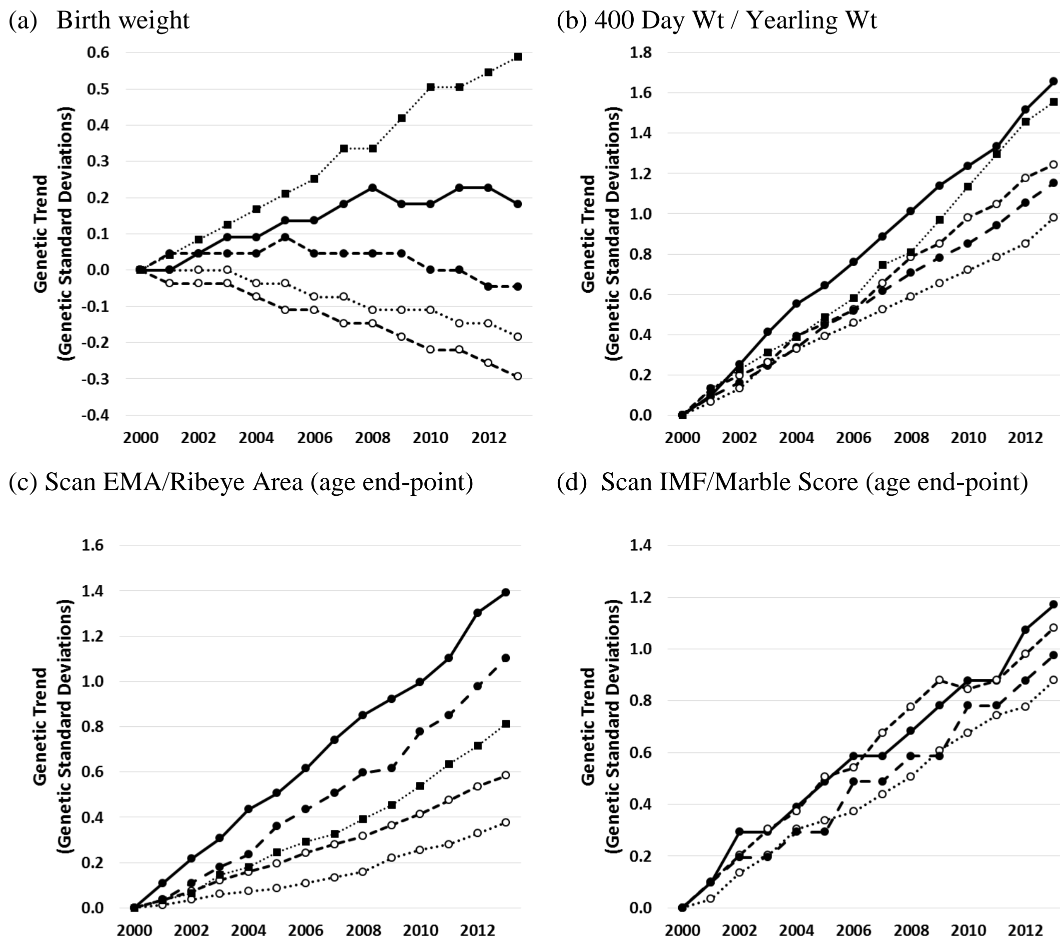


Figure 1. Standardised genetic trends for (a) Birth weight; (b) 400 Day Wt/Yearling Wt; (c) Scan EMA/Ribeye Area; and (d) Scan IMF/Marble Score in different Angus populations. Legend: —●— Australia - -○- - USA — ● New Zealand ···○··· Canada ···■··· UK

Whilst rates of genetic gain in Australian Angus have been equal to or superior to comparable beef populations, they are still well below theoretical potential gains (Johnston 2007), the rates of improvement shown in various research populations (e.g. Parnell *et al.* 1997) or those achieved in segments of the Australian sheep industry (e.g. Swan *et al.* 2009). It is suggested that the significant scope for enhanced rates of genetic improvement in the beef industry can be adequately realised within the current pipeline structure. There is a lack of evidence to indicate that re-structuring of the pipeline will have any significant impact on future rates of progress.

There is no doubt that elements of the existing beef genetic delivery pipeline can be improved to enhance rate of technology development and to address constraints to delivery of more efficient and effective recording, genetic evaluation and extension services. As suggested by Parnell (2007), there is a need for all participants in the pipeline to engage in more effective value chain partnerships to provide improved market signals, incentives, motivation and associated rewards from the application of genetics technology.

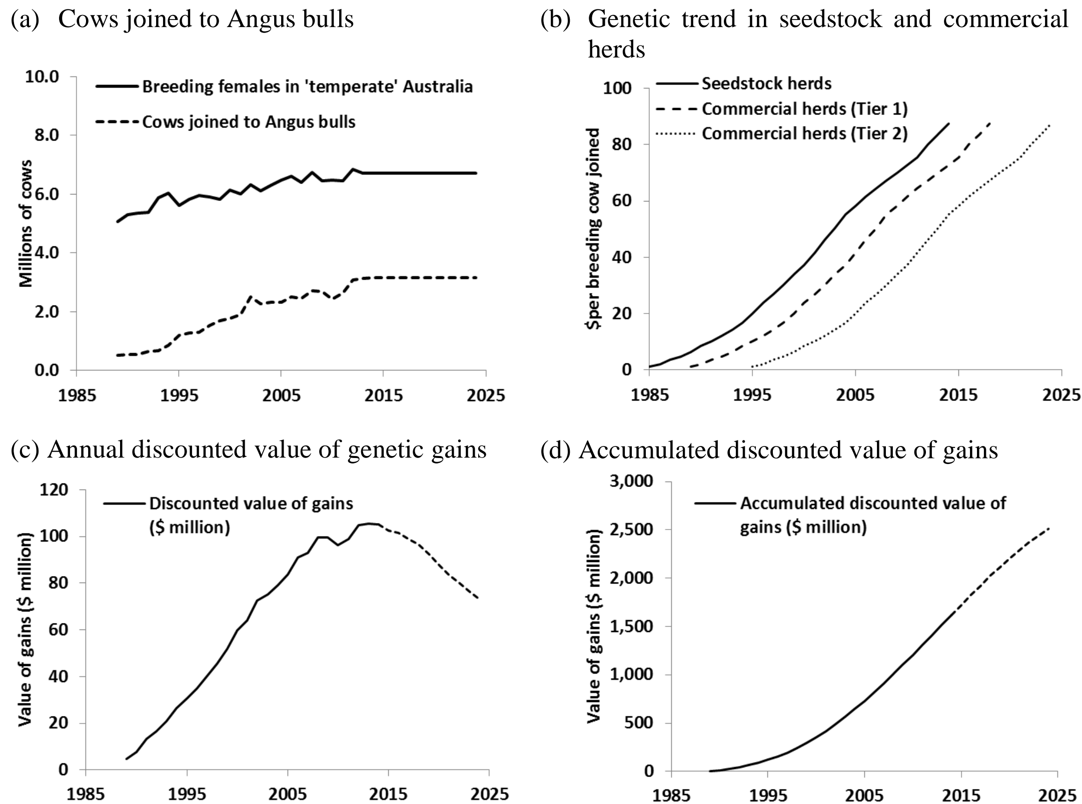


Figure 2. Modelling industry economic benefits resulting from genetic improvement in the Australian Angus population.

REFERENCES

Angus Australia (2015) http://www.angusaustralia.com.au/images/Angus_Breeding_Index.pdf

Barwick, S.A., Henzell, A.L. (2005) *Aust. J. Exp. Agric.* **45**: 923.

Farquarson, B., Griffith, G., Barwick, S., Banks, B., Holmes, B. (2003) Estimating the returns from past investment into beef cattle genetic technologies in Australia. Economic Research Report No 15. NSW Arriculture.

Graser, H-U., Tier, B., Johnston, D.J. Barwick, S.A. (2005) *Aust. J. Agric. Res.* **45**: 913.

Hammond, K. (2006) *Aust. J. Exp. Agric.* **46**: 183.

Johnston, D.J. (2007) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17**: 8.

Parnell, P.F. (2007) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17**: 167.

Parnell, P.F, Arthur, P.F., Barlow, R. (1997) *Livest. Prod. Sci.* **49**: 297.

Swan, A.A., Brown, D.J., Banks, R.G. (2009) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **18**: 326.

Woolaston, R.R (2014) MLA Final Report B.BFG.0064. Review of BREEDPLAN commercialisation model. December 2014. Meat & Livestock Australia Limited. Locked Bag 991 North Sydney NSW 2059.

Proc. Assoc. Advmt. Breed. Genet. 21: 225-228

DETAILING A BEEF GENETICS EXTENSION STRATEGY

S.J. Lee and W.S. Pitchford

School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, SA, 5371

SUMMARY

There has been a substantial increase in knowledge of the genetics underlying profit traits in Australian beef cattle. Implementation of such knowledge in breeding programs presents a significant opportunity to increase the rate of genetic gain. Current rates of genetic gain vary greatly between seedstock breeders, both within and between breeds. Commercial producers' preparedness to pay more for bulls with higher genetic merit is an important factor influencing the rate of genetic gain in seedstock herds. Meat and Livestock Australia commissioned the development of a beef genetics extension strategy focused on effective extension to various segments within the beef industry. The strategy is to focus on 1) improving knowledge amongst commercial producers to enable them to appropriately assess the value of genetic merit, thereby increasing the demand for genetically superior bulls; and 2) providing bull breeders with information to assist them in increasing the rate of genetic gain in their herds to meet the projected increased commercial demand.

INTRODUCTION

Genetic improvement allows for beef producers to increase the productivity and profitability of their enterprises and quality of their stock. The success of genetic improvement is determined by the rate of genetic improvement being achieved by seedstock enterprises and the proportion of commercial bulls that are purchased from programs achieving high rates of genetic improvement.

There have been substantial R&D efforts in beef genetics in Australia (ongoing). These have led to significant increases in understanding of the genetics underlying economically important traits and the development of DNA technology. With this knowledge and technology there is potential to significantly increase the rate of genetic gain. However, the benefit of these research outcomes is only realised when seedstock breeders utilise the technology to achieve genetic gain in a direction that will increase value chain profit. To date only a small proportion of beef genetics investment in Australia has focused on facilitating greater rates of genetic gain in the seedstock sector through adoption and effective use of BREEDPLAN. This has led to lower than optimal adoption of the technology resulting in a slower rate of genetic gain than is considered technically feasible, particularly in northern Australia (Fennessy *et al*. 2014).

Rates of genetic gain vary greatly between seedstock breeders, both within and between breeds. This is partly attributed to poor price signals from commercial producers with regard to their preparedness to pay more for bulls with higher genetic merit. Lack of commercial producer price signals is associated with ineffective proof of profit messages and a low appreciation by commercial producers of the role genetics has for enterprise productivity, product quality and profitability. The primary objective of the proposed extension strategy is to increase the rate of genetic gain and thus profitability for beef producers. Focus is given to creating demand in the commercial sector and facilitating increased rates of genetic gain in the seedstock sector.

MATERIALS AND METHODS

Over 40 stakeholders involved in beef cattle genetics extension and implementation were interviewed by the project team during the consultation phase with additional opportunity for input at a facilitated workshop for stakeholders. Stakeholders engaged in the development of the strategy included genetics extension specialists, researchers, bull breeders, commercial producers, pastoral

companies, breed society representatives and those involved in delivery of genetic evaluation and extension in Australia. The consultation focused on documenting current genetics extension and implementation efforts, identification of gaps and opportunities and exploring strategies to address the gaps and harness the opportunities. In addition to consultation, numerous industry reports of genetics implementation were considered in the development of the strategy.

RESULTS AND DISCUSSION

Six primary recommendations are detailed below.

1. Demonstration that genetics works with compelling proof of profit. Throughout the consultation there was consistent feedback on the need for the development of compelling proof of on-farm profit messages. This need was also highlighted by Freer *et al*. (2003) and more recently by Fennessy *et al*. (2014) who recommended, "Investment in generation of robust data to show the benefits of genetic improvement in commercial settings." It would be ideal to see the recommendation of Fennessy *et al.* (2014) adopted but this would be accompanied by substantial cost and a time lag to demonstration, particularly for reproductive rate. Two alternative and complementary approaches are suggested, one utilising research herd data sets and the other working closely with existing breeders who have achieved demonstrable improvement in genetic merit to demonstrate the value of genetic improvement.

Research herd data sets. It is recommended that recent research outputs be reviewed and on-farm productivity and proof of profit messages established for model farms based on differences in weaning rate, growth rate, carcass quality (and feed intake where available) that were observed for animals differing in genetic merit (teams of sires, divergent selection lines etc.). This task should be undertaken by a small team with expertise in livestock genetics, agricultural economics, livestock extension, science communication and marketing. Metrics including productivity (e.g. kg/Ha), cost of production ($/kg), turn off age, carcass quality (Meat Standards Australia Index and component traits), and return on investment (to-farm-gate value of genetic improvement) should act as a base when developing the messages and examples.

Industry case studies. The use of the industry based case studies is focused on a producer advocate approach. This approach will help facilitate the communication of messages and outcomes to commercial beef producers. These case studies will involve the development of detailed productivity and profitability outcomes through improvement in genetic merit with a longitudinal component (i.e. not once off). Case studies would ideally document the change in genetic merit achieved and associated increases in productivity (e.g. increased weaning rate, shorter time to turnoff, improved carcass quality) and income. Where possible such case studies should be undertaken in multiple regions and breeds to overcome any suggestions that the results are not applicable to particular geographic regions or breeds. Case studies should also detail the bull selection strategy employed by the seedstock enterprise to achieve the gain they have.

2. Assistance and advice to seedstock breeders. Is it expected that bull breeders new to BREEDPLAN will need assistance in understanding aspects of performance recording and genetic evaluation. Important concepts include contemporary groups, effective records, data integrity, and methods for performance recording. Whilst much of this material can be found on Southern Beef Technology Services (http://sbts.une.edu.au/, SBTS) and Tropical Beef Technology Services (http://tbts.une.edu.au/, TBTS) websites, it is essential to ensure seedstock breeders embarking on performance recording do not become disenfranchised early due to suboptimal recording methods. Support needs to be primarily targeted to seedstock breeders in northern Australia where current use of BREEDPLAN is lower than in southern Australia. Support from TBTS and local industry

service providers (on-the-ground) is likely to be required. There are currently few sufficiently experienced and available people in northern Australia to undertake this work at a local level. Through the genetics extension network (Recommendation 5) people will be identified and trained so they can fill the role of local service provision for this recommendation.

3. Influential breeder support. When investigating population structure within breed, Amer (2014) identified that approximately 60% of herds do not supply sires to other breeders and those herds that do disseminate genetics to seedstock herds tend to have higher genetic merit. Nucleus herds are defined as herds that combine superior genetic merit and high rate of genetic gain with wide dissemination of genetics. These herds should be supported to increase their progress because of the multiplier effects on the value chain.

Approach. Engagement of nucleus herds in R&D and the AGBU Influential Breeder Workshops is common practice. It is recommended that on a periodic basis (e.g. 3 years) an analysis is undertaken to identify which herds within and across breeds are the most influential with the aim of supporting current influencers (i.e. identification of nucleus herds). Two complementary approaches to engaging with these breeders are outlined:

a) Involvement in AGBU Influential Breeder Workshops: herds continue to be involved in the AGBU Influential Breeders Workshop to ensure the breeders are up-to-date with current R&D outcomes and understand how they can best utilise new technology.
b) Ensuring nucleus herds are involved in genetics R&D: Many influential herds are already involved in R&D. Where possible this should be maintained and/or expanded. There are industry benefits observed including:
   - influential breeders tend to be strong advocates for BREEDPLAN and genetic evaluation
   - they have extensive client training initiatives to highlight the benefits of genetic gain for beef producers and
   - animals in nucleus herds can inform genetic evaluation for other animals in the breed for new traits developed from R&D

4. Enhanced value chain relationships. The implementation of carcass feedback for predicted eating quality via Meat Standards Australia (MSA) Index to producers, and the use of carcass optimisation by processors provide an opportunity for clear links between improving genetic merit, increasing carcass quality and higher price received. The MSA Index is a weighted average carcass MSA score calculated from scores for individual cuts and cut weights (Watson *et al*. 2008). Recent developments in carcass optimisation allow processors to identify and segment carcasses with higher predicted eating quality so that more four and five star cuts can be marketed at substantial premiums. Market signals now exist such that carcasses with higher MSA Index attract higher price received. Two initiatives are proposed; i) continued work with beef processors to capture and value economic benefit from carcasses with higher predicted eating quality; and ii) development of tools to enable beef producers to evaluate the importance of genetic improvement on carcass quality and thus price received ($/kg). This approach is expected to facilitate clearer price signals, such that carcasses with superior predicted eating quality attract higher price received. This will provide an incentive for producers to seek bulls with superior genetic merit for eating quality and appropriately value them in addition to production traits.

5. Establish a livestock genetics extension network for training and coordination. There are many parties involved in beef genetics extension. A strong message from the consultation was the need for coordination of beef genetics extension. It is suggested that a national genetics extension network be established for the beef genetics service provider sector. The purpose of the network would be multifaceted and include:

a) Training opportunity for people involved in direct-to-producer beef genetics extension;
b) Greater awareness/coordination of the range of direct-to-producer activities occurring that have a genetic component;
c) Greater awareness of current R&D and tool development;
d) Mechanism for forming messages, providing updates and developing extension material;
e) Greater facilitation of feedback from bull breeders and commercial producers to extension agents and those undertaking R&D (a recommendation from Moreland and Hyland 2013); and
f) Planning and implementation forum.

A well-functioning genetics extension network should lead to consistent messages to industry, common extension material, and common delivery of programs. An overall aim would be that there is a high likelihood that someone in the genetics extension network would provide the same advice and recommendations to a seedstock or commercial producer for breeding program or bull selection as the next person; or identify the appropriate person for the producer to contact.

6. Market research. BREEDPLAN is an established technology that can demonstrably be used to inform animal selection and achieve genetic gain associated with greater profit for both beef producers and the wider beef value chain. Despite this, the rate of adoption and effective use of genetic evaluation to inform animal selection remains below potential. At the stakeholder workshop, there was considerable support for the engagement of a market research company to investigate barriers to adoption and to develop communication solutions to address such barriers. It is therefore recommended that a specialist market research company be engaged to investigate and report on: i) industry characteristics and barriers to adoption of genetic improvement programs at both the seedstock and commercial level; ii) key influences on decision making processes and how to leverage them; and iii) opportunities for improvement in the communication and marketing of BREEDPLAN and the economic benefits realised through genetic improvement.

Measures of success. Those consulted in the development of the strategy agreed that the success of the strategy can be measured against the following criteria listed. By 2020-
a) Performance: 50% increase in the rate of genetic progress as measured by weighted average of selection indexes for each breed society compared with 2012 base year, i.e. increase from \$4.68/year to \$7.02/year in southern Australia and \$1.04/year to >\$1.56/year in northern Australia;
b) Penetration: 25% of bulls used in commercial matings in northern Australia and 75% in southern Australia will have BREEDPLAN Estimated Breeding Values;
c) Establishment of national genetics extension and consulting network.

ACKNOWLEDGEMENTS

The authors wish to thank all those who contributed to the development of this strategy through the consultation and workshop participation. Meat and Livestock Australia funded this project.

REFERENCES

Amer P (2014) Differences among some Australian Beef Breeds and Implications for Genomic Selection. Report commissioned by MLA for Listening Team / Pipeline Consultation.

Freer R., Nicol D., Upton W. (2003) A national beef genetics extension foresight plan, A Report to MLA from the National Beef Genetics Extension Team, 2nd ed.

Fennessy P., Byrne T., Amer P., Martin G. (2014) ISBN 9781740362375.

Moreland H., Hyland P (2013) *Journal of Science Communication*. **12**: 1-17.

Watson R, Polkinghorne R, Thompson JM (2008) *Aust. J. of Exp. Agric.* **48**: 1368-1379.

Proc. Assoc. Advmt. Breed. Genet. 21: 229-232

EVALUATING DAIRY HERD GENETIC PROGRESS

M.M. Axford[1], P.W. Williams[1], D.P. Abernethy[1], and G.J. Nieuwhof[1,2]

[1]Australian Dairy Herd Improvement Scheme, 22 William Street, Melbourne, VIC 3000
[2]Biosciences Research Division, Department of Economic Development, Jobs, Transport & Resources, Centre for Agrobiosciences, Bundoora, VIC 3083

SUMMARY

The Genetic Progress Report (Report) is an effective tool for farmers and their advisors to monitor genetic change at herd level, especially when gains are only slightly positive or even negative. Farmer case studies show a willingness to refine breeding objectives based on the report. An increased uptake of the Report is expected to contribute to a more focused approach to genetic choices and increased rates of genetic gain for both profit and traits of interest.

INTRODUCTION

For over 30 years, Australian farmers have steadily achieved genetic progress for Australian Profit Ranking (APR); an economic index that includes milk, fat, protein, survival, fertility, cell count, milking speed, temperament and liveweight traits. Currently, the rate of genetic gain for Holsteins is $8.40 profit/cow/year ($11.00 and $9.04 for Jersey and Red Breeds respectively) (ADHIS, 2014). However, progress could be more than double (Lacey and Coats, 2013).

While there are likely to be many factors contributing to slower than theoretically possible genetic gain, the time between making a choice and seeing the outcome is one element. In a 2012 survey, dairy advisors who don't work in herd improvement were asked to describe characteristics of genetics that were most off-putting. The commercial politics, hard to keep up to date, time to see change, science is complicated were all reported to be off-putting by 46%, 25%, 16% and 11% respectively (n=125) (ADHIS, 2011). The results emphasise the need for unbiased, current information that is easy to access. While these findings are specific to advisors, anecdotal feedback from farmer discussion groups suggests farmers have similar needs. Simple, clear information on genetic change over time and current genetic merit would be very helpful. Furthermore such information would reinforce to farmers the permanent and cumulative nature of genetics and the need to make good choices each breeding season. The purpose of this paper is to explore the current use of a new report focussed on these aims by farmers and their advisors.

DESCRIBING AND USING THE GENETIC PROGRESS REPORT

To help farmers and advisors measure and monitor the effectiveness of their breeding choices, ADHIS developed the Genetic Progress Report. The Report is a within-breed analysis of a herd over a ten year period and shows genetic gain for APR, protein kilograms, fat kilograms, overall type, survival, daughter fertility and cell count. Over a decade's worth of herd genetic information is displayed in an easy-to-read, two page report. Parameters of the report are defined in Table 1.

The Report was first launched in April 2013 and has been released twice yearly since that time (April and August). The Report is provided to farmers, upon request, from their herd recording centre and is mailed to farmers on an annual basis.

In 2014, Australia had a total of 6,314 dairy herds with 3,023 (48%) herd recording. Of those, 2,481 herds met the minimum data requirements of the Report in August 2014. The proportion of total reports by breed was 82% Holstein (or Holstein cross), 15% Jersey (or Jersey cross), 3% other breeds. Each of the 2,481 herds will have a different Report and reactions to the report will vary. The following case studies provide an overview of two approaches.

Table 1. Definitions of parameters included in the Genetic Progress Report

ABV	Australian Breeding Value, Equivalent to EBV. ABVs for protein kg, fat kg, overall type, survival, daughter fertility and cell count are reported
APR	Australian Profit Ranking, Australia's economic index prior to March 2015.
Average APR	Average APR of current cows in the herd.
Average of top 10%	The average APR of cows in herds ranked within the top 10% of the breed by year of birth.
Breed	Purebred and non-purebred are considered side by side, but other breeds separately.
Current cows	Cows calved in the past 30 months (relative to release date) and contributed data to an ABV.
Genetic trend for each trait and index	Increased = Average APR or ABV of years 6-10 > Average of years 1-5 and the last APR or ABV > first APR or ABV Decreased = Average APR or ABV of years 6-10 < Average of years 1-5 and the last APR or ABV < first APR or ABV No clear trend = if either of the above statements are not true
GBG	Percentage of cows with sires included in the Good Bulls Guide or Progeny Test near the time of their dam's mating.
Minimum data requirement	At least 50 cows of a single breed with ABVs. Dates of birth occur over several years.
National Average	The average APR or ABV of cows of the same breed and year of birth.
Rank	Rank within breed. Sorted by APR then ASI.

In April 2015 ADHIS launched three new breeding indices. As a result the GBG has been updated replacing the APR with the new economic index, Balanced Performance Index (BPI). The Health Weighted Index (HWI) and Type Weighted Index (TWI) are also included.

Case Study 1 – Chris and Diana Place. Finding out their Holstein herd ranked in the top 100 for APR was a welcome surprise to Chris and Diana Place, but they were more interested to see opportunities to improve fertility and mastitis resistance through breeding. These are just some of the results from the herd's Genetic Progress Report. Chris dairies with his wife, Diana and his brother Peter, in Western Victoria. Their 420 cow Holstein herd averages more than 285 kg fat and 235 kg protein from a predominantly grass-based feeding system, 5% above the regional average of herd recorded cows (ADHIS, 2014).

Breeding decisions have always focused on high production cows that are easy to milk and have few health and fertility problems. For many years Chris has selected bulls from the top of the list for APR and within that list, bulls that are positive for udders, feet and legs. Their Report, of which sections are shown in Figure 1, shows how much has been achieved with this consistent approach. While it's reassuring to see how much can be achieved through breeding, Chris was more interested in the sections of the Report that showed opportunities for improvement. Figure 1 showed that genetic progress for fertility is declining so Chris immediately gave higher priority in his selection decisions.

When it comes to selecting sires for the season, Chris normally uses the Good Bulls Guide. His strategy is to go straight to the top four or five APR bulls in the Guide and then check them for the individual traits that are important for his herd. So from now on, he plans to look at fertility and mastitis resistance as well udders, feet and legs. 'The Report presents our herd data in a very useful format. It's a great tool to help us with our breeding decisions.'

Figure 1. Genetic Progress Report Place herd (Protein kg, Fat kg, Cell count and Fertility)

Case Study 2 – Bettina and John McLeod. Bettina and John McLeod's herd ranks in the top 150 for profit among Holsteins as shown in Figure 2 but the McLeods do not spend hours studying bull catalogues. Their achievement of a high genetic merit herd has come through a successful partnership with their breeding advisor, Mr Graeme Heaver.

The McLeods dairy in South-West Victoria. Their 400 cow herd averages 770kg milk solids/cow, 56% greater than the regional average (ADHIS, 2014). A couple of times a year Mrs McLeod and Mr Heaver discuss the herd's breeding objective, progress towards that objective, and specific priorities for the coming joining season.

The Report has been useful in fine tuning the McLeod's selection criteria. "The report highlighted how much progress has been made for cell count in recent years. Satisfied with this result, Bettina and I decided to place higher priority on selecting for fertility within the top sires" Mr Heaver said.

Mrs McLeod finds the Report particularly useful for identifying and learning from past mistakes. "When I look at the graphs, I'm interested to see the dips – because they show where we made a mistake." For example in 2006, another breeding advisor chose the sires, purely for type

without consideration of other traits. While the report shows a subsequent peak in genetic progress for type, it was accompanied by dips in profit, fat, protein, fertility, and longevity. Mrs McLeod is keen to be continually improving their dairy operation and uses the Report as a tool for monitoring breeding progress. “The Report presents our data in graphs that make it easy to see long term trends,” she said.

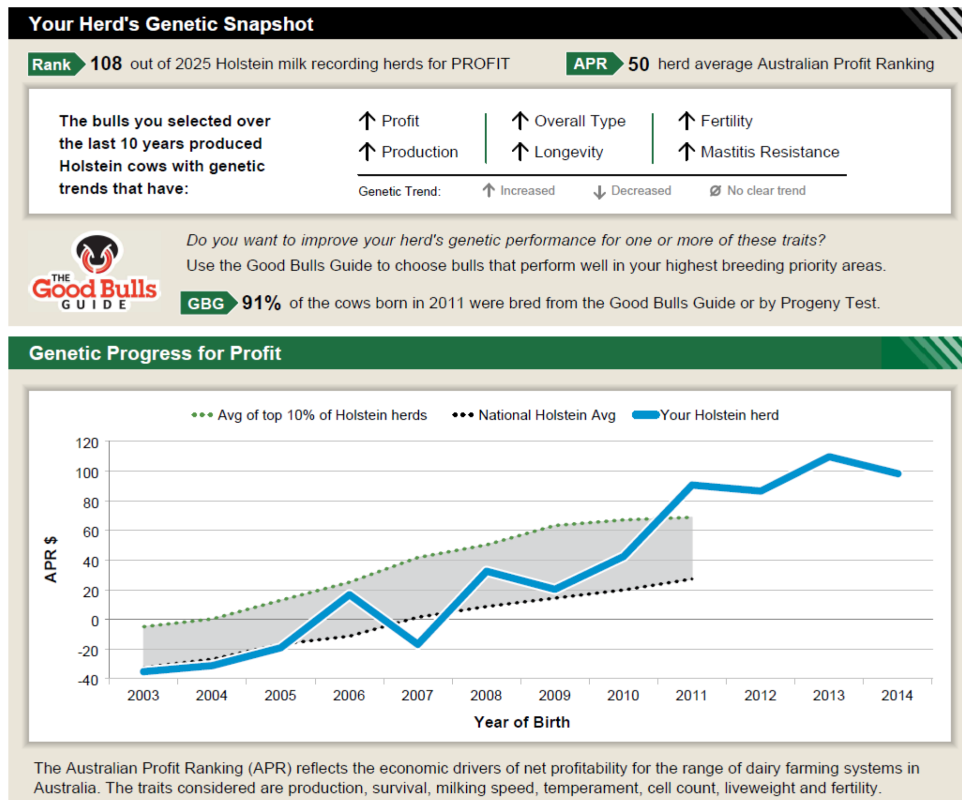


Figure 2. Genetic Progress Report for John and Bettina McLeod’s herd (Profit)

CONCLUSIONS

The Genetic Progress Report is a practical output of ADHIS and Dairy Australia’s investment in genetic evaluation, genomic technology and genetics extension. Its value in identifying success and opportunities for improvement is shared amongst farmers and their professional advisors. Further work is scheduled to comprehensively evaluate the success of the Report. Updates to the report to incorporate Australia’s new breeding indices will further enhance opportunities to use this Report in genetics extension and education activities.

REFERENCES

ADHIS (2011) Good Bulls Guide – Dairy Advisor Survey 30 May 2011 (unpublished).

ADHIS (2014) Australian Dairy Herd Improvement Report 2014.

Lacey R. and Coates S. (2013) Development of genetics scenarios and implications for the Australian dairy industry. Final Report, submitted to Dairy Australia, April 2013, commercial and in confidence.

Proc. Assoc. Advmt. Breed. Genet. 21: 233-236

REALISING GENETIC IMPROVEMENT FOR THE EXTENSIVE LIVESTOCK INDUSTRIES AS A WHOLE

R. G. Banks
Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

SUMMARY
This paper examines some aspects of the overall performance of livestock improvement systems, first asking what we mean by the term "system". Variation in behaviour of agents within the system is examined, and some tentative conclusions about the nature of such systems, and scope for their management, proposed.

INTRODUCTION
In discussions of livestock genetic improvement, one theme focusses on improvement at a whole of industry or country level (eg. Hill, 1971; Smith, 1978; Hammond, 2006), and the term "system" or something similar is sometimes used. In general such discussion is descriptive, retrospective and focussed on averages, often leading to reasoning in support of various forms of collective action. In this context, "collective" means activities funded, initiated, managed etc. via mechanisms or agencies acting on behalf of often large numbers of individual enterprises.

A frequent observation is that rates of genetic improvement are less than what is technically possible, and that this can be attributed at least partly to various forms of market failure negatively impacting individuals' incentives to invest in recording and to select somehow optimally. In turn, this thinking has underpinned various forms of collective investment into aspects of livestock genetic improvement in many countries (Amer et al, 2012).

This paper is an attempt is made to explore the nature of such systems, and to ask whether taking a "systems perspective" can help achieve some different outcomes.

This general perspective rests on some important assumptions, including:

- Agents (individuals or organisations within the system) will respond to the availability of knowledge, tools etc by adopting and implementing them relatively immediately and uniformly
- Interactions or interdependencies between tools and/or agents are either minimal or benign in their effects (which ironically is not unlike assuming that all genetic variation is additive, or that at least we can ignore interactions)
- Information flow is such that all agents have perfect information about the future in a form relevant to whatever decisions they are able to make.

These are all heroic assumptions. Where this heroism is recognised, the usual response is to propose some form of collective or even government intervention, which usually takes the form of subsidising the cost of some system input(s). A common example is that government or industry funds are used for R&D and E, which in essence means making some of the raw materials of genetic improvement – knowledge, tools and skills – either free or cheap.

Embedded in these assumptions is the expectation (hope?) that market returns will be sufficient to fund the risky investments that for example sire breeders must make.

We can identify some aspects of the system through simple questions.

a) What comprises the system?

In the livestock industries, the components that are interacting include: bull- and ram-breeders in breeds within the breeding sector, commercial producers, finishers, processors, retailers, service

* AGBU is a joint venture of NSW Department of Primary Industries and University of New England

providers to these sectors (including research and development organisations), and consumers.

b) How do agents interact?

Agents interact via the sale of packages of genes (mainly sires) and their derivatives (commercial stock and their products). There are many sellers and buyers in most livestock system markets. What do prices for genetic material tell us about information flow in beef and sheep?

Data relevant to this question is incomplete. However, Van Eeenanam (2012) analysed sales data from a number of Angus studs, and found clear relationships between price paid and bulls' merit, with r-squared in the 15-35% range, regression coefficients for the indexes in use at time of \$40-70 per index point when stud bulls sales are excluded. Walkom (pers. comm.) analysed sales data from a number of ram breeding flocks, and found r-squared values in the range 35-45% and regression coefficients of price on index of \$45-120. Analysis of Angus herd average prices indicates an r-squared of 45% for herd average bull price on herd average \$Index merit.

So, some information is being exchanged and used in at least some transactions. With more data, a greater understanding of the variation in this exchange, whether it is changing over time, and longer term trends in market share could be examined.

c) Are the agents all aiming for the same things?

It is possible to examine variation in breeders' attitudes and behaviour around breeding objectives. A survey approach, such as in the MERINOSELECT Breeding Goals Survey (MLA, 2012) reveals more variation in the weighting proposed for individual traits amongst breeders within a wool type category than between categories. A more analytical approach, involving calculating correlations between response vectors amongst breeders, reveals similar variation in realised breeding direction in Angus cattle (Lee, 2014) and Merino sheep (Swan, pers. comm.)

QUESTIONS ABOUT LIVESTOCK IMPROVEMENT SYSTEMS

This brief overview suggests that there is a system of interacting and diverse agents. What questions can be asked of such systems? Smith (1978) is an example of studies highlighting the fact of differing perspectives within such systems – the question really being "what are agents trying to achieve?"

- differences between individual breeders and the national or industry
- differences between different sectors in an industry

In most countries and industries, the general response to these differences has been the establishment of some form of national or industry-level evaluation, with either mandatory or voluntary participation. Such evaluation, especially when based on BLUP methods, has lead or contributed to, increases in rates of genetic progress. Depending on the interaction between the information generated by the evaluation and aspects like bull registration, such initiatives can effectively override differences in perspectives, but this may come at some political and financial cost. Amer *et al* (2012) examined returns from genetic improvement in beef cattle in several countries, and identified a range of approaches differing widely in apparent cost per animal, but less so in outcome.

The general question of aims leads inevitably to the specific question of breeding objectives and hence indexes. National evaluation systems such as BREEDPLAN, Sheep Genetics and ADHIS invest significant resources into developing "industry" objectives, usually with inputs from breeders and others in industry, but may also provide scope for customised indexes. As breeders' confidence in breeding technologies grows, the situation can arise, and seems to be arising more frequently, where some breeders wish to innovate by making and demonstrating genetic progress for new traits. By definition this leads to a situation where only some breeders have information on the new traits, which in turn means that the less innovative breeders are

automatically at a disadvantage if indexes are modified to include the new traits.

This question of different views on objectives (and hence on indexes) in turn leads to the question of how to value outcomes of the system. If there is some national objective, but a significant portion of industry is pursuing objectives correlated at less than 1 with the national objective, should the national objective be used to value improvement? And the question of how to value outcomes leads to the question "how to improve those outcomes?"

RESPONDING TO GAPS – LEAVE IT TO THE MARKET?

One way of viewing the establishment of industry-wide genetic evaluation systems is that they are primarily meant to provide information, enabling a market in genetic material to function with greater efficiency. This outcome seems to develop, but slowly and unevenly (Van Eenennaam, 2012). This outcome alone provides a strong basis for some form of collective investment: objective information available to buyers and sellers is necessary for an efficient market. Such information is also a necessary condition for genetic improvement, but is not sufficient on its own. In the absence of direct payments for genetic improvement itself, Amer (pers. comm.) has noted that the primary incentive for breeders to make genetic progress is possible future market share. This incentive would appear to be the driver for dairy breeding companies, who operate in competitive markets in multiple countries, and pig breeders (Knap, 2014). It is not obvious whether beef and sheep breeding businesses in Australia are sufficiently scaleable to make growing market share a realistic goal for all but a minority of enterprises.

Another form of investment in information is via the calibration of genomic tests. This service is provided in the US dairy industry (and hence effectively, globally) by the USDA, and in the Australian beef and dairy industries by the respective industry funding bodies. This initiative has the beneficial effect of enabling informed investment in genetic information products, but raises the interesting question of who pays for the performance data required for the calibration. If such information is collected for other purposes, as is the case for some traits in dairy, this may partly obviate the problem, but where this is not the case, funding data collection becomes yet another dilemma poised around "who benefits, who pays?"

These are two forms of collective intervention, but on their own, they may not generate much overall improvement. What might we need to understand in order to do better?

RESEARCHABLE QUESTIONS – FIRST ORDER

The fundamental level of research must always be focussed on analysis of genetic variation, including new methods. Having accurate descriptors of animals' genetic merit is absolutely essential for efficient selection and the efficient operation of the market for genes (or the animals that carry and transmit them).

The next level of analysis is where we start to understand the behaviour of the "system" and its components. Industry databases now offer scope to analyse what individual enterprises have done – including tracking selection differentials, accuracies achieved, levels of linkage and resulting flows of information between enterprises, and directions of selection. Tools have been developed for such analysis, such as Takestock (Johnston, pers. comm.), but as yet they are not being used systematically to help design interventions (such as new services and products).

Reverter *et. al* (2011) provided an example of preliminary analysis of the network properties of the meat sheep genetic improvement system, and suggested further steps. Similarly, Charteris *et al* (2001) suggested the use of agent-based modelling to explore both the properties of livestock improvement systems, and potentially to explore through simulation the effects of different interventions.

RESEARCHABLE QUESTIONS – DESIGN AND COORDINATION

The bigger questions are around issues such as:

- If we can identify information gaps, should they be filled, and if there is a cost, who pays and how?
- If we assume that all system behaviours ultimately depend on decisions at the individual enterprise level, can we "nudge" them in any way (Thaler *et al*, 2008). For example, if relying on market premiums to incentivise bull- and ram-breeders, and we know that those premiums are imperfect signals, should someone add some sort of rewards for "good behaviour" – things like additional recording, sampling new bloodlines, strengthening linkage, or even simply making faster progress?
- And ultimately, who monitors system behaviour and develops responses? This is ad hoc in Australia, relying on varying inputs and capacities of organisations such as MLA, breeds, ABRI, previously CRCs, and breeders, each with their own constraints. The problem is that this ad hoc system appears to be associated with opportunity costs running into hundreds of millions over the medium term.

CONCLUSIONS

Charteris *et al* (2001) proposed that livestock improvement systems could usefully be examined using the perspective of Complex Adaptive Systems (CAS), where a CAS exhibits "strong interactions among its elements so that current events influence the probability of many kinds of later events". The brief discussion here suggests that there is merit in that recommendation. Already it is clear that there is considerable variation at the level of individual enterprise (agent in CAS language) behaviour and at least some inter-agent interactions.

It is not so clear how to tackle system-wide coordination and management (research using the CAS framework might generate useful ideas for this), but it seems highly likely that in the absence of any such coordination that overall performance will fall short of what seems possible (although what we think is possible may reflect ignorance on important inherent limits of such systems). Responding through some sort of central control seems politically unlikely, so the search must continue for ways to do better. Simply piling up raw materials (new knowledge), ensuring accurate EBVs and Indexes and other types of current R&D may unfortunately be necessary but not sufficient.

REFERENCES

Amer, P.R., Banks, R.G., and Garrick, D.J. (2012) *ICAR 38*
http://www.icar.org/cork_2012/Abstracts/Published/Amer.htm

Charteris, P.L., Golden, B.L., and Garrick, D.J. (2001) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **14**: 461.

Hammond, K. (2006) *Aust. J. Exp.Agric*. **46**:183.

Hill. W.G. (1971) *Animal Production* **13:** 37.

Knap, P. (2014) *Proc. WCGALP* **10**

Lee, S., Nuberg, I..K., and Pitchford, W.R. (2014) *Proc. WCGALP* **10**

Reverter, A., Dominik, S., Field, S., and Banks, R.G. (2015) *Proc. Ass. Advmt, Anim. Breed. Genet.* **19**:243.

Smith,C. (1978) *Animal Production* **26**:101.

Thaler, R.H., and Sunstein, C. R. (2008) *Nudge* Yale University Press

Van Eenannam, A.L. (2012) *ICAR 38*
http://www.icar.org/cork_2012/Abstracts/Published/Van%20Eenennaam.htm

Proc. Assoc. Advmt. Breed. Genet. 21: 237-240

INHERITANCE OF TAIL LENGTH IN MERINO SHEEP

J.C. Greeff, L.J.E. Karlsson and A.C. Schlink

Department of Agriculture and Food Western Australia, 3 Baron Hay Court Perth, WA, 6051

SUMMARY

The inheritance of tail length and spine length was estimated using data of 2667 Merino lambs assessed at approximately 6 weeks of age on the marking cradle. Maternal permanent environmental effects affected spine length and marking weight but not tail length. The heritability of tail length and spine length was both 0.58 ± 0.05, while the genetic correlation between the two traits amounted to 0.58 ± 0.05. However, adjusting the data for marking weight or spine length removed the genetic correlation between spine and tail length. The heritability of a subjective score for tail length was 0.38 ± 0.05. The results indicate that selection for short tails is possible and that it will not have a negative impact on spine length provided adjustment is made for body weight or spine length.

INTRODUCTION

Breech and tail strike are the most common types of blowfly strike suffered by Merino sheep. Mulesing and tail docking are therefore used to reduce the impact of the main predisposing factors such as wrinkles, dags and breech cover that contributes to breech strike. Mulesing removes the skin around the anus, and docking the tail at the 3 or 4th joint, makes animals less susceptible to breech and tail strikes because it reduces the accumulation of faecal material and urine in the breech and on the tail (James, 2005).

Recent trends in animal welfare and ethical sheep production systems, question these surgical techniques. This has resulted in alternative methods being investigated to remove or reduce the impact of the predisposing factors to breech strike. Greeff *et al.* (2013) and Smith *et al.* (2009) have shown that dags, wrinkles, urine stain and high breech cover scores are the most important indicator traits and selecting against these traits will reduce breech strike. Watts *et al*. (1977) showed that tail length played an important role in determining the susceptibility of sheep with diarrhoea to breech strike. Sheep with very short tails are more susceptible to breech strike than longer tails because they cannot lift their tail to hold the wool out of the way when defaecating and urinating. James (2005) therefore suggested that breeding for shorter tails should be considered in un-mulesed sheep to make sheep less susceptibility to breech strike.

However, breeding for short tails may result in skeletal abnormalities as was found by James *et al.* (1990; 1991) in Merino sheep where single dominant genes were the mode of inheritance. Shelton (1977) showed that tail length adjusted for body length had a heritability of 0.38 in Rambouillet sheep. Scobie and O'Connell (2002) showed that the mode of inheritance of tail length in different sheep breeds was additive. However, no study has estimated the genetic correlation between tail length and spinal length. This study was carried out to determine whether it would be possible to breed for short tails, and whether there is any negative relationship between tail length, spinal length and body weight in Merino sheep.

MATERIAL AND METHODS

Animals. Body weight, tail length, spinal length and a visual score for tail length were recorded on 2667 lambs that were the progeny of 62 sires mated to 1294 ewes and born from 2012 to 2014 in the Australian Wool Innovation Breech strike flock at Mt Barker research station in Western Australia. Lambs were born over approximately 6 weeks from mid July to end of August

every year. Full pedigrees, sex of the lamb, birth status and age of the dam were recorded on all lambs.

Measurements. Tail length was measured at marking at approximately 6 weeks after birth. A tape measure was permanently fixed length-wise in the marking cradle. The lamb was placed in the cradle with its spine lying lengthwise on the tape. A measurement was taken at the joint between the skull and the first neck vertebrae of the lamb, at the root of the tail, and at the tip of the tail. Spinal length and tail length for each lamb were calculated through subtraction. Tail length was also scored by holding the lamb's leg perpendicular to its body and laying the tail along the anterior side of the backleg over the hock. Lambs were scored from 1 (short) to 5 (long) depending on the length of the tail relative to the hock. Tails that touched the hock were given a score of 3, while shorter tails were given scores of 1 or 2, and longer tails 4 or 5 depending on length. Any lamb with evidence of their tails being bitten off, were not recorded.

Body weight of each lamb was recorded at marking in 2013 and 2014. As marking weight was not recorded in 2012, a body weight at marking was estimated for the 2012 drop by multiplying the average daily gain from birth to weaning at 85 days of age, with the average age of the 2012 drop at marking and adding birth weight. This method assumes that growth was linear which may not have been the case. However, it was deemed acceptable in a preliminary study such as this until more data are recorded.

Data analysis. The data were analysed with ASREML (Gilmour *et al.* 2009). An animal model with and without maternal permanent maternal environmental effects was fitted with year of birth (3 years), sex (male or female), age of the dam (2 to 6 years) and birth status (single or multiples) as fixed factors and all 2 way interactions. Day of birth was fitted as a covariate within year of birth. Maternal pedigrees were not fitted because of shallow pedigrees. Different univariate analyses were first carried out with and without body weight at marking as a covariate to identify significant fixed effects. The following combinations of direct additive and maternal permanent environmental effects were fitted.

$$y = Xb + Za + e \qquad (1)$$
$$y = Xb + Za + Wpe + e \qquad (2)$$

where y, b, a, pe and e are the vectors of observed traits of animals, fixed effects, direct additive genetic effects, permanent maternal environmental effect and residual effects, respectively. X, Z and W are incidence matrixes for fixed, direct additive genetic and permanent maternal environmental effects of y, respectively. Marking weight was also fitted as a covariate to tail and spine length, while tail length was also adjusted for spine length to determine its impact on the inheritance of tail length. Log likelihood ratio tests were carried out amongst the models to determine the most appropriate model for each trait. This was followed by bivariate analyses between tail length, spinal length and body weight to obtain variances and covariances for genetic parameter estimation, by fitting the most appropriate model as determined by the previous analysis. Tail score was only analysed to estimate the correlations with tail length.

RESULTS AND DISCUSSION

Table 1 shows the average spine length, tail length, tail score and marking weight at marking at approximately 6 weeks of age.

Year of birth, sex of the lamb, birth status, day of birth and age of the dam affected tail length, tail score, spine length and marking weight significantly (P<0.001). No significant interaction effects were found between these fixed effects. Log likelihood ratio tests shows that model 2

which included both direct additive genetic and permanent maternal environment effects, fitted the spine and marking weight data best while model 1 with only direct additive genetic effects fitted the tail length and tail score data best.

Table 1. Average spine length, tail length and body weight at marking (approximately 6 weeks of age).

Trait	n	Mean	SD	CV	Min	Max
Spine length (cm)	2665	73.4	6.02	8,2	46	98
Tail length (cm)	2665	23.2	3.62	15.6	11	40
Tail score	2661	3.7	0.67	18.1	1	5
Marking weight* (kg)	2665	13.0	3.09	23.8	4.2	25.4

* Include some estimates

Fitting marking weight as a covariate to spine length removed all the permanent maternal environmental effects and resulted in model 1 fitting the data best. The variance components are shown in Table 2.

Table 2. Phenotypic variation, heritability and permanent maternal environmental effects of tail length, spine length, body weight at marking, and tail score, and fitting marking weight or spine length as covariate to tail length and tail score.

Parameter	Tail length	SE	Spine length	SE	Marking weight	SE	Tail score	SE
Vp	9.5		31.5		7.7		0.40	
h^2	0.58	0.05	0.58	0.05	0.44	0.06	0.38	0.05
h^2	0.48[a]	0.05[a]	0.51[a]	0.04[a]	---	---	0.36[a]	0.05[a]
h^2	0.54[b]	0.05[b]	---	---	---	---	0.35[b]	0.05[b]
m^2_{pe}	---	---	0.06	0.02	0.17	0.03	---	---

Vp = Total phenotypic variation with model 1; h^2 = direct additive heritability; m^2_{pe} = maternal environmental effect, [a] Marking weight fitted as covariate; [b] spine length fitted as covariate

Heritability estimates. Tail and spine length and tail score were all heritable traits with tail and spine length having the highest heritability of 0.58 followed by an estimate of 0.44 for marking weight and 0.38 for tail score. Maternal permanent environmental effects were not significant for tail length but it made a significant contribution ($P<0.01$) to marking weight and for spine length. The heritability estimate of 0.58 for Merinos is higher than the heritability of 0.39 in Rambouillet sheep (Shelton, 1977). However, it is not clear whether Shelton (1977) used the measured tail length, or a tail length adjusted for body length in his analysis. Fitting marking weight as covariate in this study decreased the heritability of tail length from 0.58 to 0.48 and decreased the heritability of spine length from 0.58 to 0.51. When tail length was adjusted for spine length, the heritability of tail length decreased slightly to 0.54. Scobie (2002) reported a very high heritability of 0.82 for tail length which is of the same magnitude as the estimate of 0.77 reported by Branford Oltenacu and Boylan (1974). However, both these studies worked with crossbred sheep in which the short-tail Finnish Landrace featured prominently. In the more common type of sheep breeds, Branford Oltenacu and Boylan (1974) reported an estimate of 0.50 which is slightly lower than the estimates derived in this study.

Correlations. Table 3 shows the phenotypic and genetic correlations between tail length, spine length, body weight at marking and tail score. Tail length was phenotypically positively correlated with spine length (0.44) and with body weight (0.53) at marking. Similarly, moderately strong genetic correlations were found between tail length and spine length (0.58) and between tail length and marking weight (0.67). However, fitting marking weight as covariate removed the strong correlation between tail and spine length. This indicates that tail and spine length are independent traits and that the genetic correlation is induced through body weight. Tail score was genetically moderately strongly correlated with tail length. Although the heritability of tail score is less than that of tail length, and has a correlation with tail length that is lower than expected, it may still be a useful trait to select indirectly for short tails without resorting to direct measurements.

Table 3. Phenotypic (above diagonal) and genetic (below diagonal) correlations between tail length, spine length and marking weight and their standard errors in brackets.

	Tail length	Spine length	Spine length [a]	Marking weight	Tail score
Tail length		0.44 (0.02)	0.05 (0.02)	0.53 (0.02)	0.61 (0.02)
Spine length	0.58 (0.05)		----	0.81 (0.01)	----
Spine length [a]	-0.01 (0.08)	----		----	----
Marking weight	0.67 (0.02)	0.80 (0.03)	----		----
Tail score	0.77 (0.05)	----	----	----	

[a] Fitting marking weight as covariate

CONCLUSIONS

This study shows that tail length is a heritable trait and that it would respond to selection. It has a moderately strongly genetic correlation with body weight at marking but adjustment for body weight or spine length at marking removed the genetic relationship between tail length and spine length. This indicates that Merino breeders can breed for shorter tails without having any negative impact on spine length provided adjustment is made for body weight or body size.

ACKNOWLEDGEMENTS

Australian Wool Innovations Ltd. is thanked for funding this project.

REFERENCES

Branford Oltenacu, E.A.B. and Boylan W.J. (1974) J. Heredity **65**: 331.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) ASREML User Guide Release 3.0. VSN International Ltd., Hemel Hempstead. HP1 1ES, UK. www.vsni.co.uk.

Greeff J.C., Karlsson L.J.E. and Schlink A.C. (2013) Anim. Prod. Sci. **54**: 125.

James P.J. (2005) Aust. J. Exp. Agric. **46**: 1.

James P.J., Gare D.W., Singh A.W., Clark J.P., Ponzoni R.W. and Ancell P.M. (1990) Wool Tech Sheep Breed. **38**: 106.

James P.J., Ponzoni R.W., Gare D.R. and Cockrum K.S. (1991) Proc. Assoc. Advmt. Anim. Breed. Genet. **9**: 404.

Scobie D.R. and O'Connell D. (2002) Proc. NZ Soc. Anim. Prod. **62**: 195.

Shelton M. (1977) J. Heredity **68**: 128.

Smith J., Brewer H. and Dyall T. (2009) Proc. Assoc. Advmt. Anim. Breed. Genet. **18**: 334.

Watts J.E. and Marchant W.D.A. (1977) Aust. Vet. J. **53**: 118.

Proc. Assoc. Advmt. Breed. Genet. 21: 241-244

TESTING AN APPROACH TO ACCOUNT FOR DAUGHTER MISIDENTIFICATION IN THE ESTIMATION OF BREEDING VALUES

N. Lopez-Villalobos[1], J.R. Bryant[2], B.L Harris[3] and D.J. Garrick[1,4]

[1] Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand
[2] DairyNZ Limited, Hamilton, New Zealand
[3] Livestock Improvement Corporation, Hamilton, New Zealand
[4] Department of Animal Science, Iowa State University, Ames, IA, USA

SUMMARY

A stochastic simulation was developed to account for daughter misidentification (DaMi) in the estimation of breeding values (EBV) of bulls. Miss-assignment of parentage was simulated in 20% of daughters of 50% of bulls in 30 of 40 commercial herds. A bias of -40 kg protein was assigned to those cows that were misidentified as heifer calves. The model for genetic evaluation included fixed effects of contemporary group and random effects of additive animal, permanent environment and residual. The approach to account for DaMi was to include a regression coefficient for each sire that reflected DaMi. Compared to progeny test (PT) EBVs, parent average (PA) EBVs were overestimated by 2.3 and 3.4 kg protein for bulls with and without DaMi, respectively. Compared to PT EBVs, reproof (RP) EBVs were underestimated by 3.5 kg and overestimated by 5.1 kg protein for bulls with and without DaMi, respectively. The model that accounted for sires with DaMi removed biases from PA to PT for both groups of bulls, but compared to PT EBVs, RP EBVs were underestimated by 1.0 kg protein for both groups of bulls. This set of simulations indicates that paternity-verified status for a sire-daughter record can remove biases in genetic evaluation caused by DaMi.

INTRODUCTION

Studies in New Zealand dairy cattle (Johnson and Harris, 2010) have documented that estimated breeding values (EBV) of young bulls based on parent average (PA) are typically higher than subsequent EBV which include daughter information from the progeny test (PT) or from reproof (RP) in commercial herds. These biases from PA to PT and from PA to RP breeding values are not unique to New Zealand dairy bulls; there is evidence of similar biases in Denmark (Pedersen *et al*. 1995), the United States of America (Powell *et al*. 2004) and Germany (Rensing *et al*. 2009). Possible factors causing these biases include preferential treatment of bull dams (Pedersen *et al*. 1995) preferential treatment of daughters (Kuhn and Freeman, 1995), misidentification of sire-daughter pairs (Winkelman, 2013), heterogeneous variance across herds and years and effects of heterosis (van der Werf *et al*. 1994).

Using DNA parentage verification the magnitude of sire to daughter misidentification in commercial herds has been estimated at 23% (Bowley et al., 2012). However in PT herds, where sires are initially evaluated for widespread commercial use, the degree of sire misidentification was approximately 5% in Livestock Improvement Corporation PT herds for the 2005 to 2007 seasons (Winkelman, 2013). The impact of daughter misidentification on dairy sire breeding value estimation has been investigated in New Zealand (Winkelman, 2013). Estimates of progeny group yield means for milk, fat and protein of DNA-verified daughters were higher than those of daughters for which paternity had been assigned via mating records. Progeny of genetically superior sires can easily be assigned to inferior sires and their genetic evaluations are biased downward (Bowley et al., 2012). The underestimation of elite sires has been shown via simulation

to have a negative impact on rates of genetic gain in the dairy cattle population (Bowley *et al.*, 2012).

In this study a stochastic simulation was developed to evaluate biases caused by daughter-sire misidentification and to test an approach to correct for this bias during successive predictions of breeding values of bulls from birth to 9 years of age representing PA, PT and RP.

MATERIALS AND METHODS

Simulation scheme. The cow population comprised 40 commercial herds and another 20 herds for the sire proving scheme (SPS). Average herd size was 518 lactating cows plus replacements. Protein yield (Py) for each lactation of a cow was calculated as:

$$Py = \mu + TBV + year + herd + p + e$$

where μ is the general mean of the population (μ = 160 kg protein); TBV is the true breeding value of an animal modelled as TBV = 0.5×(TBVsire + TBVdam) + (z x 0.7 x genetic SD) where z is a random number from a normal distribution with mean=0 and SD=1; year is a year effect simulated from a normal distribution with mean 0 and variance=14.0 kg^2; herd is a herd effect simulated from a normal distribution with mean 0 and variance=28.0 kg^2; p is a permanent environmental effect of the cow through her productive life and repeated for each lactation; e is a residual effect for each lactation. Estimates of genetic, permanent and residual variances were 82.1 kg^2, 85.5 kg^2 and 111.7 kg^2, respectively.

Genetic evaluation. Genetic evaluation for protein yield was performed each year using an animal model. The model equation for genetic evaluation was the following:

$$y_{ikm} = \mu + hyl_i + a_k + p_m + e_{ikm}$$

where y_{ikm} is the protein yield record for cow m in contemporary group i defined as cows of the same lactation number calving in the same herd and year (hyl), a_k is the random additive genetic effect of animal k, p_m is the random permanent environmental effect of cow m, and e_{ikm} is the random residual effect unique to y_{ikm}. Lactation yields of protein were mature equivalent.

The best 68 cows and best 6 progeny-tested bulls were used to produce 34 young bulls to be progeny tested in SPS herds, the best 20% of these were selected when the bulls were 5 year old based on progeny test results. Selected bulls were used in commercial herds for two years.

Simulation of daughter-sire misidentification and bias. A bias of -40 kg protein was added to the yields of 20% of cows in 30 of 40 commercial herds. The introduction of this bias was at the level of contemporary group (herd-year-lactation number) and the bias was repeated for the same cows across all repeated lactations. This introduction of bias was to represent miss assignment of parentage whereby genetically elite sires get assigned progeny of inferior sires. Those cows were progeny of 50% of the bulls selected at random. All progeny of the other 50% of the bulls were correctly assigned to their sire. The model equation for genetic evaluation to attempt to remove bias caused by daughter misidentification was the following:

$$y_{ijkm} = \mu + hyl_i + \beta_j + a_k + p_m + e_{ijkm}$$

where y_{ijkm} is the protein yield record for cow m in contemporary group i, daughter of sire j and β_j is the fixed regression coefficient for sire j that reflects misidentification (0 if rightly assigned daughter and 1 if wrongly-assigned daughter).

RESULTS AND DISCUSSION

Table 1 shows distributions of true and estimated breeding values for protein yield and their reliabilities for 238 bulls evaluated at different ages; 129 bulls had 20% of daughter

misidentification in each of 30 commercial herds and 109 bulls were without daughter misidentification. Compared to PT EBVs, PA EBVs were overestimated by 2.3 kg protein for bulls with daughter misidentification and 3.4 kg protein for bulls without daughter misidentification. When selected bulls entered the bull team and a significant number of daughters in the commercial population contributed to the reproof of these bulls, a divergent bias occurred for the two groups of bulls. Compared to PT EBVs, RP EBVs of bulls with daughter misidentification were underestimated by 3.5 kg protein whereas RP EBVs of bulls without daughter misidentification were overestimated by 5.1 kg protein.

The biases in protein EBVs from PA to PT in this simulation agree with Johnson and Harris (2010) for New Zealand dairy cattle. They reported overestimation of 3 to 4 kg protein of PA EBV compared to PT EBV, and suggested that such bias could reflect drift in genetic evaluations (Powell *et al*. 2004) and may include parentage error associated with sires of sons as they accumulate daughter numbers.

Protein EBVs from progeny test results were similar for both bull groups, but RP EBVs of bulls without daughter misidentification increased whereas RP EBVs of bulls with daughter misidentification decreased. This trend was expected in the simulation because misidentification of sire-daughter was always penalised with a negative bias. This agrees with Winkelman (2013) who reported that EBVs for production traits were, on average, biased downwards when all progeny was not DNA-verified.

The model that accounted for sires with misidentified daughters, on average, removed biases from PA to PT for both groups of bulls, but compared to PT EBVs, RP EBVs were underestimated by 1.0 and 1.2 kg protein for bulls with and without daughter misidentification, respectively

Livestock Improvement Corporation has offered SNP-based DNA sire verification services to customers since mid-1990s. The test is based on approximately 100 SNPs, where recorded sire was deemed correct if concordance with his daughter was at least 99% (Winkleman 2013). Likewise, customers of CRV Ambreed can obtain DNA sire verification services via Genomz. Test results from either service provider are recorded in the national database and nine classes of verified paternity of cows can be derived, including DNA-verified and paternity assigned using mating and calving records. Further research is being undertaken to include parentage verification status in the genetic evaluation for each sire for all traits as an attempt to correct biases caused by daughter misidentification.

Table 1. Distributions of true (TBV) and estimated (EBV) breeding values for protein yield and their reliabilities for 129 bulls that had 20% of misidentified daughters and 109 bulls without misidentified daughters, evaluated in different years following selection on protein EBV obtained with *a model without adjustment* for daughter misidentification

		TBV		EBV		Reliability	
Age of bull	N	Mean	SD	Mean	SD	Mean	SD
		Bulls with 20% of misidentified daughters					
0	129	16.4	3.6	18.8	7.5	36	1.8
4	129	16.4	3.6	16.5	6.0	83	1.6
9	129	16.4	3.6	13.0	7.4	89	3.7
		Bulls without misidentified daughters					
0	109	16.7	3.9	20.0	7.0	36	1.6
4	109	16.7	3.9	16.6	6.4	83	1.7
9	109	16.7	3.9	21.7	12.0	89	4.0

Table 2. Distributions of true (TBV) and estimated (EBV) breeding values for protein yield and their reliabilities for 120 bulls that had 20% of misidentified daughters and 118 bulls without misidentified daughters, evaluated in different years following selection on protein EBV obtained with *a model that included adjustment* for daughter misidentification

		TBV		EBV		Reliability	
Age of bull	N	Mean	SD	Mean	SD	Mean	SD
Bulls with 20% of misidentified daughters							
0	120	18.0	4.6	20.2	3.9	36	1.7
4	120	18.0	4.6	20.2	5.5	83	1.5
9	120	18.0	4.6	19.2	5.0	89	3.1
Bulls without misidentified daughters							
0	118	18.1	4.8	20.6	3.7	36	1.9
4	118	18.1	4.8	20.6	5.5	83	1.5
9	118	18.1	4.8	19.4	5.3	90	4.0

CONCLUSION

This simulation study shows that misidentification of sire-daughter pairs can be a source of bias in the genetic evaluation of dairy sires, a model that includes parentage verification status in the genetic evaluation for each sire can potentially correct for this bias.

REFERENCES

Bowley F.E., Amer P.R. and Meier S. (2012) *Proc. 33rd Conf. Int. Soc. Anim. Genet.* P1004.
Johnson D.L. and Harris B.L. (2010) *Proc. 9th World Cong. Genet. App. Livest. Prod.* Leipzig, Germany. Communication No. 0227.
Kuhn M.T and Freeman A.E. (1995) *J. Dairy Sci.* **78**: 2067.
Pedersen G.A., Christensen L.G. and Petersen P.H. (1995) *Acta Agric. Scand. Sect. A, Anim. Sci.* **45**: 26.
Powell R.L., Sanders A.H. and Norman H.D. (2004) *J. Dairy Sci.* **87**: 2614.
Rensing S., Pasman E. and Reinhardt F. (2009) *Interbull Bulletin* No 40 pp. 123.
Van der Werf J.H.J., Meuwissen T.H.E. and de Jong G. (1994) *J. Dairy Sci.* **77**: 3174.
Winkelman A.M. (2013) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **20**: 25.

Proc. Assoc. Advmt. Breed. Genet. 21: 245-248

MERINO BREEDING OBJECTIVES UNDER CLIMATE CHANGE

P. Graham[1], A.A. Swan[2] and R.G. Banks[2]

[1]NSW Department of Primary Industries, Yass, NSW 2582
[2]Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

SUMMARY

Breeding objectives were constructed for Merino enterprises operating in environments subject to forecast climate changes to 2030. These were derived using gross margins per dry sheep equivalent (DSE) from the GrassGro pasture model at three locations in New South Wales, Yass, Woolbrook, and Narrandera, and two enterprises, wool and dual purpose. Gross margins were predicted to be lower under forecast climate change, particularly at Narrandera. Simple breeding objectives were derived using GrassGro data, and selection index predictions based on these showed that losses in profitability could mostly be offset, and improved. The GrassGro breeding objectives were strongly correlated with those currently used in the MERINOSELECT genetic evaluation system, although there were differences in outcomes for body weight.

INTRODUCTION

The potential impact on future pasture production from increased temperatures and a more variable rainfall pattern has been assessed by the Southern Livestock Adaption 2030 project (sla2030.net.au). Even when the predictions are for minor changes in rainfall, increases in temperature result in a decrease in soil moisture leading to a decline in pasture production, both quantity and quality. Combined with a predicted increase in the number and severity of dry events, breeding flocks will likely have to adjust stocking rates downwards to manage the increased pressure on the pasture base, or utilise more conserved feed. If these adjustments are not made there will be an increase in the loss of perennial species from the system. The reduced use of fertiliser through the high rainfall zone of Australia adds to the pressure from climate change. Under these conditions it is relevant to ask what the appropriate breeding directions for livestock production are. In this paper, we use the GrassGro pasture model (Moore *et. al.* 1997) to estimate gross margins for Merino sheep enterprises under predicted climatic conditions. We then derive simple breeding objectives targeting these climate scenarios and compare them to the indexes currently used by the Merino industry

MATERIALS AND METHODS

GrassGro enterprise modelling: GrassGro was used to model pasture production of annual and perennial species in three locations in New South Wales, using local soil and daily weather records from 1960 to 2013. The locations were Yass (stocking rate between 1960 and 2013 of 14.2 DSE/ha), Woolbrook (13 DSE/ha), and Narrandera (4.7 DSE/ha). Pasture production to 2030 was then modelled using climate predictions of temperature and rainfall from the HadGEM2 Global Circulation Model (Jones *et al.* 2013) with stocking rates set with a restriction to maintain a minimum acceptable ground cover. These were 10 DSE/ha at Yass, 9.3 DSE/ha at Woolbrook, and 3.1 DSE/ha at Narrandera. Two different Merino enterprises were compared at each location, a wool system and a dual purpose system with all ewes mated to terminal sires and lambs finished in a feedlot. Production means for ewe fleece weight (agfw, kg), fibre diameter (afd, microns), body weight (awt, kg), and number of lambs weaned (nlw, lambs per ewe joined) are shown in Table 1.

* AGBU is a joint venture of NSW Dept. of Primary Industry and the University of New England

Gross margins ($/DSE) were calculated within GrassGro based on these production means, and median price and cost figures from 2009 to 2013.

Table 1: Trait means for Wool and Dual Purpose (DP) enterprises at Yass, Woolbrook and Narrandera

Trait	Yass		Woolbook		Narrandera	
	Wool	DP	Wool	DP	Wool	DP
agfw (kg)	5.3	5.5	3.2	5.5	6.0	5.5
afd (μ)	18.6	19.5	17.3	19.5	20.0	19.5
awt (kg)	52	55	47	55	58	55
nlw (lambs)	1.02	0.98	0.84	1.01	1.08	1.08

Breeding objectives for future pasture production: Starting from these base scenarios, relative economic values were calculated for each trait by independently increasing the mean by 5% and re-running GrassGro to obtain a new gross margin. These were deviated from the base gross margin and converted to a unit change for each trait (on the scale of MERINOSELECT breeding values) to obtain final relative economic values.

Responses to selection on indexes for these breeding objectives were predicted over a 15 year time horizon for a ram breeding program with 300 ewes mated to 10 rams annually, selection intensities of 0.78 and 1.99 in females and males, and generation intervals of 4.5 and 3 in males and females. Two sets of selection criteria were compared: base traits including fleece weight, fibre diameter, CV of diameter, and body weight, and base traits plus number of lambs weaned.

Comparison with modified MERINOSELECT objectives: MERINOSELECT provides standard indexes for three breeding objectives: Dual Purpose (DP+), based on a meat-focussed production system where surplus progeny are sold as lambs and a portion of ewes are joined to terminal sires; Merino Production (MP+) based on a balanced wool and meat production system where surplus progeny are sold as hoggets; and Fibre Production (FP+) based on a wool focussed production system. Importantly, the objectives feature reducing emphasis on body weight from DP+ to MP+ to FP+. They also differ in the level of emphasis placed on fibre diameter. DP+ is designed to increase fleece weight while maintaining fibre diameter, MP+ balances increased fleece weight and reduced fibre diameter, while FP+ targets large reductions in fibre diameter while holding fleece weight constant.

Genetic correlations between GrassGro and MERINOSELECT breeding objectives were calculated as $a_1'Ca_2/\sqrt{(a_1'Ca_1.a_2'Ca_2)}$, where a_1 and a_2 are vectors of relative economic values for GrassGro and MERINOSELECT objectives and C is the genetic covariance matrix between traits in the objectives. The MERINOSELECT objectives include traits which cannot be modelled by GrassGro, including carcass traits in DP+, worm egg count in FP+, and staple strength in all three. Economic values for these traits were not included when calculating correlations. The comparison is therefore based on modified MERINOSELECT breeding objectives containing equivalent traits to those modelled by GrassGro.

RESULTS AND DISCUSSION

Gross margins per DSE for the base production system parameters are shown in Table 2 for the period 1960 – 2013, and in 2030 under predicted climate changes. Profitability was predicted to be lower under forecast climate change for all locations and production systems, but particularly at Narrandera. Also shown are gross margins for 2030 where each trait mean was independently changed by 5%. For fleece weight, fibre diameter, and number of lambs weaned, the trait changes always led to an increase in profitability. By contrast, increasing body weight by 5% reduced profitability in wool systems at all three locations, and for the dual purpose system at Woolbrook.

For dual purpose systems at the other sites, increasing body weight had a neutral effect on profitability at Yass, and a positive effect at Narrandera. Note that gross margins were considerably higher for dual purpose systems because it was only possible to model a system where all ewes were mated to terminal sires, and the results do not incorporate the cost of replacing the ewe flock.

Table 2: Gross margins ($/DSE) for base production systems from 1960 to 2013, with predicted climate changes in 2030, and 2030 with each trait mean changed by 5%.

Period/Trait	Yass		Woolbrook		Narrandera	
	Wool	DP	Wool	DP	Wool	DP
			Gross margins ($/DSE)			
1960 – 2013	19.93	28.75	15.69	30.22	16.17	22.60
2030	18.60	28.16	14.41	29.81	9.36	13.45
+ 5% agfw	20.00	29.52	15.81	31.07	10.65	14.41
- 5% afd	19.90	29.90	17.31	31.17	10.65	15.17
+ 5% awt	18.00	28.19	13.87	29.63	8.33	15.17
+ 5% nlw	19.20	29.03	14.95	30.38	10.33	15.17

Predicted trait and gross margin responses from index selection on the GrassGro-derived objectives are shown in Table 3. Improvements in gross margin ranged from $3.58 to $6.88 per DSE, and were large enough to offset and improve on the loss in profitability predicted due to climate change at Yass and Woolbrook, but not at Narrandera. For wool traits, most systems resulted in balanced improvement of fleece weight and fibre diameter, the exceptions being the wool system at Woolbrook in which there was a greater emphasis on fibre diameter, and the dual purpose system at Narrandera in which a large increase in body weight limited the gain in fleece weight. There were large reductions in body weight for all wool systems (-2.14 to -2.93kg), consistent with negative economic values in these systems. In dual purpose systems there were smaller reductions at Woolbrook (-0.70 to -0.81kg), little change at Yass (0.07 to 0.27kg), and large increases at Narrandera (4.11 to 4.85kg), once again consistent with the respective economic values. When base traits only were recorded and included in the selection index, number of lambs weaned decreased for all systems except dual purpose at Narrandera. By contrast, when number of lambs weaned was recorded, changes in the trait were either neutral or positive for all systems (0.0 to 0.08 lambs), and gross margin was increased by small to moderate amounts.

Table 3: Trait and gross margin (GM, $/DSE, in the predicted 2030 environments) responses after 15 years of selection with base and base+nlw selection criteria

Criteria	Trait	Yass		Woolbrook		Narrandera	
		Wool	DP	Wool	DP	Wool	DP
base	agfw	8.29	6.45	1.34	8.19	6.97	1.04
	afd	-0.92	-1.19	-1.77	-1.00	-1.04	-0.62
	awt	-2.14	0.27	-2.14	-0.70	-2.93	4.85
	nlw	-0.03	-0.01	-0.02	-0.02	-0.04	0.04
	profit	4.06	4.05	6.82	3.58	3.78	5.73
+nlw	agfw	7.95	6.03	1.37	7.85	6.20	1.06
	afd	-0.86	-1.08	-1.74	-0.93	-0.88	-0.53
	awt	-2.18	0.07	-2.19	-0.81	-2.78	4.11
	nlw	0.00	0.03	0.00	0.01	0.02	0.08
	profit	4.21	4.32	6.88	3.72	4.23	6.42

Genetic correlations between GrassGro and the reduced MERINOSELECT breeding objectives (Table 4) generally show a high degree of association, indicating that the industry indexes will select sheep that increase profit under the predicted climatic conditions. For wool systems, the FP+ objective showed the highest correlations, and this was because it has the lowest level of emphasis on body weight and is therefore best aligned with objectives which have a negative emphasis on the trait. The MERINOSELECT objectives were more highly correlated with dual purpose systems, again because of the (increased) level of emphasis on body weight in these. The DP+ objective was most highly correlated with the dual purpose system at Narrandera. It is clear from these results that the key difference between the two approaches (GrassGro and MERINOSELECT) is their treatment of the impact of body weight on profitability. In the GrassGro model, body weight is considered as a single trait, while in the MERINOSELECT model body weight is separated into yearling and mature body weight, with the former targeting improved growth rates in sale lambs. The MERINOSELECT model does account for the cost of feed, and the outcome of this is that economic values are typically positive for yearling weight and negative for mature weight, but because genetic correlations between ages are strongly positive overall genetic gains for body weight are most often positive. We also note that because the GrassGro and MERINOSELECT objectives model feed in different ways, it may be difficult to resolve the differences in the treatment of body weight.

Table 4: Genetic correlations between GrassGro and reduced MERINOSELECT breeding objectives

Objective	Yass		Woolbrook		Narrandera	
	Wool	DP	Wool	DP	Wool	DP
DP+	0.76	0.91	0.55	0.82	0.85	0.94
MP+	0.82	0.94	0.67	0.89	0.86	0.89
FP+	0.89	0.97	0.87	0.94	0.88	0.82

CONCLUSIONS

The pasture modelling presented here shows that the future profitability of sheep enterprises will be lower under predicted climate change, but that in some locations genetic improvement programs can offset and to a degree exceed the loss in profit. Current MERINOSELECT breeding objectives are relatively well aligned with the genetic change required, but further work is required to understand the effects of body weight, and in particular to fully understand the biological and economic effects of reducing body weight at some or all stages of the productive lifetime of animals.

REFERENCES

Jones, C.D., Hughes, J.K., Bellouin, N. *et. al.* (2011) *Geosci. Model Dev.* **4**: 543.

Moore, A.D., Donelly, J.R. and Freer, M. (1997) *Agr. Syst.* **55**: 535.

Proc. Assoc. Advmt. Breed. Genet. 21: 249-252

SNP-BASED PARENTAGE IN AN AUSTRALIAN CATTLE INDUSTRIES CONTEXT: DOES ONE SIZE FIT ALL?

R.E. Lyons[1], S. Buttsworth[1], D. Waine[1] and M. Kelly[2]

[1]The University of Queensland, School of Veterinary Sciences, Gatton, QLD 4343, Australia.
[2]The University of Queensland, Centre for Animal Science, Brisbane, QLD 4062, Australia.

SUMMARY

Globally, there is a trend away from microsatellites or short tandem repeats (STRs) to single nucleotide polymorphisms (SNPs) on the basis of perceived advantages for genetic identification, traceability and assessment of parentage. However, the transition is not simple and here we highlight unique problems faced in adapting low cost SNP-based assays for Australian systems.

INTRODUCTION

Much has been written about the promises of SNP-based parentage verification in livestock and animal traceability across the supply chain (Heaton *et al*. 2002, Van Eenennaam *et al*. 2007, Baruch and Weller 2008). Advantages discussed includes abundance, amenability to high-throughput genotyping platforms and reproducibility across laboratories. The biallelic nature of the marker along with automation of bioinformatic analysis makes this process less prone to reader error. Unlike microsatellites, they are synergistic with existing genomics applications and hence more cost-effective for those breeds undertaking genomics-based breeding programs now or in the near future. Based upon modelling data and validation in taurine breeds the International Society of Animals Genetics (ISAG) recommended a set of 100 core SNP, and later added an additional set of 100 markers to increase the exclusion power in indicine and synthetic breeds.

Following development and optimisation of Sequenom SNP panels, we demonstrate that the ISAG-recommended core bovine SNP parentage panel is not sufficient to provide accurate parentage verification in many common Australia production systems. The objectives of this study are: (1) to demonstrate factors influencing effectiveness of the tests, (2) develop additional analyses to clearly identify, communicate and eliminate problems pre- and post-analysis, and (3) maximise accuracy and completeness of parentage verifications especially in large test cohorts.

MATERIALS AND METHODS

Samples and DNA Extraction. Commercial populations of Brahman or Brahman-cross animals were used as case studies for sire verification only. DNA was extracted and purified from hair follicles using customised protocols.

Genotyping. Genotyping was performed using iPLEX reagents and platinum protocols for high multiplex PCR, single base primer extension (SBE) and generation of mass spectra, as per the manufacturer's instructions (Sequenom, San Diego). SEQ1 iPLEX panels contained a total of 138 SNP including 95 ISAG core plus 4 ISAG additional SNP. The additional panel in SEQ2 consisted of 59 SNPs for a combined total of 197 markers genotyped and total of 97 ISAG core SNP. These new markers were developed to be informative in Brahman and Tropical Composite breeds. Mass spectra were analysed using TYPER software (Sequenom, San Diego) in order to generate genotype calls and allele frequencies. Some sires were genotyped using the custom GeneSeek Genomic Profiler low-density BeadChip (GGP-LD) with ~ 25,000 SNPs assayed per sample.

Post-genotyping data analysis. To identify issues of mislabelling or sampling errors, duplicate sample checks were performed by counting the number of discordant marker calls between two samples. Less than or equal to 5 discordant markers between genotypes across assays were considered likely to be from the same animal, indicative of sampling or testing errors and

requiring recollection to ensure accurate genotypes. Prior to requesting recollection of samples, putative duplicates were routinely checked via microsatellite analysis to ensure these samples did not represent closely related individuals (e.g. full sib). To date all cases have been confirmed as the same individual (n=19). Hence a discordance threshold of ≤ 5 SNP to represent potential duplicate samples appears a suitable value.

Populations and Primary Analysis. Batches represented small, medium and large multi-sire matings, hence denoting increasing degrees of complexity in sire assignment. The small batch contained 20 Brahman progeny and 5 sire candidates. The medium-sized batch included 173 Brahman crosses with 26 sire candidates. The large batch had 706 Brahmans originating from 3 properties with a total of 42 sire candidates (Table 1). Each batch was initially parent verified using the SEQ1 SNP data. Parentage analysis was via exclusion based on opposing homozygotes with strict criteria (exclusions ≤ 3). In the large herd, the SEQ2 SNP test was used to assess for increased accuracy of parentage assignment (exclusions ≤ 3). Any sire-progeny matches with a misconcordance rate > 3 SNP were not accepted.

Detection of potential sibs. Often in large extensive beef herds it is not possible for the breeder to supply all potential sires. Thus we tested the ability of SEQ2 to assign unqualified animals to sib groups for a set of 204 Brahman progeny with known sire information. This represented progeny of 29 sires with an average of 7 progeny per sire (min=1 and max=17). These animals representing a subset of the large multi-sire population discussed previously. The accuracy of detecting the sib families was assessed against the known sire to group potential sibs from a method using genomic relationship matrix (GRM) developed with SEQ1 and SEQ2 panels. The GRM matrix was formed using all animals in the genotyped parentage analysis as per VanRaden (2008). The subset of animals requiring allocation to sib families was then selected from within this full matrix. Potential sib groups were formed by successively adding animals to the sib group if their mean relationship with the current group of the new animal was greater than an empirically defined threshold. In the current analysis this threshold was varied from 0.12 through to 0.2 to examine the trade off in accuracy and number of animals assigned. In practice the threshold could be determined by analysis of animals with known parentage within the same parentage population.

RESULTS AND DISCUSSION

Broadly speaking, the process for the provision of large multi sire herd parentage analyses is as follows: owner provides hair samples to laboratory with a list of offspring and potential parents, lab staff prepare and genotype samples, data is analysed and results are returned to owner. The measure of success of a parentage verification is the proportion of all calves correctly assigned with no resubmission of samples required. However this is often not the case with a number animals remaining unresolved following the initial analysis. Unresolved cases may be due to incorrect sample submission, unrepresented sires/dams in the analysis, or the inclusion of genotyping errors with less than acceptable call rates. Data on true batches of increasing complexity are shown in Table 1 which illustrates a number of important considerations in deciding upon the test panel chosen.

The small commercial batch had 100% of progeny assigned to a sire. All sires are accounted for and present in the testing pool. For the medium sized commercial herd 82% of available progeny were matched to a sire despite 2 sires and 12 progeny failing to reach an acceptable SNP count (n ≥ 120 or ~ 85% of the total markers). These were unavailable for retesting and not included in the primary analysis. Upon consultation with the owner, one additional sire candidate was identified and genotyped. When this sire was included in a reanalysis, the overall total increased to 89% of available progeny matched to a sire which is comparable to that achieved through microsatellites.

Table 1. Case studies from commercial batches across small, medium and large populations

POPULATION	BREED	TEST	PROGENY (n)	SIRES (n)	Assigned (%)	Unassigned (%)	Recollect (%)
SMALL	Brahman	SEQ1	20	5	100	0	0
MEDIUM: analysis 1	Brahman X	SEQ1	173	24	82	11	7
MEDIUM: analysis 2	Brahman X	SEQ1	173	25	89	4	7
LARGE:Analysis 1	Brahman	SEQ1	706	42	57	35	8
LARGE:Analysis 2	Brahman	SEQ2	706	42	97	3	0

As shown in Table 1, the rates of successful assignment are affected by a number of factors including size and completeness of the animal data set provided at initial testing. To better identify the factors that may be leading to failures in assigning parents affecting results, pre- and post-analysis data assessment tools have been generated to identify and resolve issues in a timely fashion. For example, the duplicate genotype check before parentage verification has been invaluable in identifying problems such as transcriptional errors during sampling or laboratory error, and as such saves time and increases accuracy/confidence as recollects for suspect samples can be organised promptly. Similarly, and especially for large batches of sample, graphic representations such as that shown in Figure 1 can be helpful in demonstrating issues and corrective steps required to resolve the analyses.

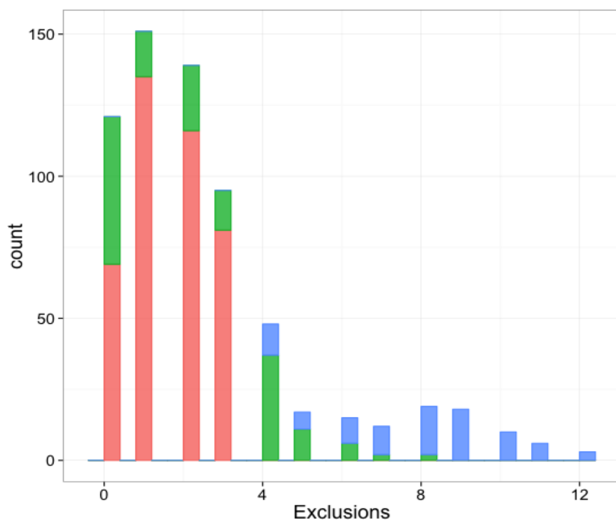


Figure 1. Representation of data quality and reasons for unassigned progeny in initial testing. Pink represents qualified animals, Green represents those requiring retesting, and Blue represents missing sires.

Noting the increased complexity of the large commercial batch, it would be recommended to future clients that all sires are genotyped using the GGP-LD test with progeny and dams on the lower density SEQ1/SEQ2 assay. Generally the number of sires is small in comparison to the total number of animals requiring genotyping and the marginal cost of genotyping the sires on a higher density panel is low and provides three immediate benefits: firstly the SNP array tends to be more accurate, secondly they have higher call rates and thirdly any particularly difficult to resolve cases (progeny) can be upgraded and testing across a much larger set of markers. While there is significant benefit in upgrading sires in particular to GGP-LD, the economic impost of genotyping all animals (sires and progeny) on this platform remains an impediment, and genotyping progeny on the smaller assays offsets this cost. Previous analyses have shown the importance of increasing marker counts in these large herds, as shown in the primary analysis in Table 1 where only a subset of markers representative of the SEQ1 test panel are used in parentage verification. While 97% of progeny matched to a sire using SEQ2 data for progeny (analysis 2), had the SEQ1 test option been chosen only 57% of progeny were resolved.

It is important to note that the 97% assignment for analysis 2 was achieved only after extensive consultation with owners. In the first iteration of the SEQ2 analysis, 79% of progeny were assigned to a sire which is still superior to that of the SEQ1 test.

With large multi-sire groups where herds may have been across multiple properties or extensively grazed, the nature of the enterprise makes complete mustering and collection of all sires logistically and economically difficult and it is not uncommon to have missing sires. While missing sires are evident through numbers of unassigned progeny, it is not immediately clear if one or more sires are missing. To this end, the use of GRM has been investigated to cluster progeny and hopefully give an estimate of missing sire numbers. *In silico* testing using CRC data for progeny (n = 204) was used to identify an appropriate stringency to allow clustering whist retaining relative accuracies. A minimum mean sib relationship of 0.13 was deemed the best balance between accuracy and percentage assigned. In practice this threshold would be determined by examining the mean relationship in animals within each data set where parentage is known. It is important to note that this is not intended to be a verification tool, but rather used as a guide only to estimate the likely number of sires not submitted and the likely groups of sibs. The benefits of returning estimated sire-groups for unresolved calf cases are considerable. Using this data and specifically animal groupings, producers may be able to identify the missing sires by paddock, birth/joining date or even physical characteristics of the calf groups.

Table 2. Accuracy and assignment rates of GRM analyses for clustering unassigned progeny in parentage verification analyses

	Stringency				
	0.12	**0.13**	**0.14**	**0.15**	**0.2**
Accuracy	85%	93%	93%	94%	100%
% assigned	60%	55%	50%	41%	19%

While there are very good arguments for the shift from microsatellites to SNP-based parentage and in the long term SNP will become the dominant mode of identification in parentage and traceability testing, the difficulties associated with this transition have often been understated and largely unreported. The Australian cattle industry with its large diversity of breeds and extensive production systems represents a sector where unique challenges to successful transition exist. Previous research has demonstrated that ISAG's core (100 SNP) panel can be inadequate for parentage testing for some breeds (Strucken et al. 2014) and our recent experiences suggest this is also the case for large *Bos indicus* cohorts in Australian production systems. Clear communication between laboratory and producers is essential including identification of challenges or constraints to achieving high levels of assignment. For example, tight deadlines for verification on consanguineous groups can be best met if sires and/or dams are collected and genotyped in advance of the progeny. Correct sampling techniques (clean and adequate amounts) is important. Improved results obtained over three seasons for a large Brahman herd from 71% assignment in year 1 and 2, to 89% in Year 3 were largely attributable to improved sampling technique on-farm. It is our view that implementation of these additional analyses at strategic points across the pipeline will further enhance rates of assignment.

REFERENCES

Baruch, E. and J. I. Weller (2008) *Animal Genetics* **39**: 474.
Heaton, M. P., G. P. Harhay, G. L. Bennett, *et al*. (2002) *Mammalian Genome* **13**: 272.
Strucken, E. M., B. Gudex, M. H. Ferdosi, *et al*. (2014) *Animal Genetics* **45**: 572.
Van Eenennaam, A. L., R. L. Weaber, D. J. Drake, *et al*. (2007) *J. Anim. Sci.* **85**: 3159.
VanRaden, P. M. (2008) *J. Dairy Sci.* **91**: 4414.

Proc. Assoc. Advmt. Breed. Genet. 21: 253-256

IMPACT OF SCANNING LEAN CATTLE ON THE GENETIC CORRELATION BETWEEN SCAN AND CARCASE INTRAMUSCULAR FAT IN ANGUS AND HEREFORD CATTLE

S.F.Walkom, M.G.Jeyaruban and D.J.Johnston

Animal Genetics and Breeding Unit[1], University of New England, Armidale, NSW 2351, Australia

SUMMARY

BREEDPLAN, the Australian beef cattle genetic evaluation system, uses ultrasound scan intramuscular fat as a correlated trait for predicting carcase intramuscular fat. More recently, it has been observed that seedstock herds are being scanned at younger ages and lower levels of fatness and this research was undertaken to examine the effect on heritability and the genetic correlation estimates when scan records are removed using fat depth thresholds. Using BREEDPLAN data to estimate these genetic relationships, this study yielded genetic correlation estimates of 0.37 and 0.36 in Angus and 0.69 and 0.54 in Hereford for bull and heifer scan intramuscular fat, respectively. The results showed a useful improvement in the genetic correlation between bull intramuscular fat and carcase intramuscular fat in Angus cattle to 0.48. However, for Angus heifers and Hereford bulls and heifers there was no significant improvement, suggesting that strategies to reduce lean scanning will not improve the genetic correlation estimates in those cases.

INTRODUCTION

Current genetic correlation estimates used in BREEDPLAN, the Australian beef cattle genetic evaluation system, (Johnston *et al.* 1999) between scan intramuscular fat (IMF) and abattoir carcase intramuscular fat (CIMF) are based on a pooled temperate breeds analysis from the Beef Cooperative Research Center I (Reverter *et al.* 2000). Subsequent re-analysis of the genetic correlations between scan intramuscular fat and CIMF of industry data have produced estimates for Angus (Reverter and Johnston 2001; Bŏrner *et al.* 2013) and Hereford (Reverter and Johnston 2001; Meyer *et al.* 2004) lower than the pooled breed analysis.

Compared to 2010 the scanning of Angus bulls in 2014 occurred on average 40 days younger due to a trend towards producers scanning bulls at 400-days. Bŏrner *et al.* (2013) estimated that scanning younger bulls (mean age 426 days) compared to older bulls (590 days) reduced the genetic correlation for scan intramuscular fat and CIMF from 0.43 to 0.34. It was hypothesised that the genetic variation in intramuscular fat was not being expressed at the younger age, or could not be detected by the ultrasound machines in the leaner cattle.

Preliminary analysis (pers. comm. M.G. Jeyaruban) showed that the lower genetic correlations were, in part, a consequence of scanning a large number of contemporary groups (CG) with a mean rib fat below 3mm and a mean P8 fat below 5mm. However, BREEDPLAN users were cautious of implementing rib and P8 fat restrictions to scan IMF records as restrictions would lead to a large proportion of records, up to 50% in Angus bulls, being excluded from evaluations, unfairly disadvantaging herds with genetically lower fat levels.

The objective of this study was to determine the merit of using rib and P8 fat depth thresholds as criteria to improve the genetic correlation between scan IMF and CIMF in Angus and Hereford breeds. The study also explored the impact of retaining lean CG that had large variation in scan IMF on genetic correlation estimates.

[1]AGBU is a joint venture of the NSW Department of Primary Industries and University of New England

MATERIALS AND METHODS

The study analysed data submitted to the Angus Society of Australia and Herefords Australia databases for BREEDPLAN evaluations prior to November 2014. The majority of CIMF records in the Angus and Hereford data sets were from the Beef CRC (Reverter *et al.* 2000) with only a few breeders progeny testing and recording CIMF. 7,833 Angus CIMF and 1,836 Hereford CIMF records were used in this study (Table 1 and 2). The Angus ultrasound scanning data set contained 226,687 BIMF and 245,840 HIMF records (Table 1) and there were 86,603 BIMF and 70,020 HIMF records for Hereford (Table 2).

The following approaches were taken to remove lean CG prior to estimation of the genetic correlations between scan and carcase intramuscular fat.

- All data: all scan intramuscular fat records were retained
- Subset 1: Scan CG with a mean rib fat depth below 3mm and a mean P8 fat depth below 5mm are removed except for CG with a mean P8 > 4mm and where the sum of the CG's mean and standard deviation for P8 fat was greater than 5
- Subset 2: sub set 1 + CG with high phenotypic variation for scan IMF (sd. of IMF in top 25%)

For both the Angus and Hereford heifers the analysis was repeated with a more stringent cut off based on a rib fat depth of 5mm and P8 fat depth of 7mm.

Statistical Analysis. Genetic variances, correlations and variance ratios were estimated by applying restricted maximum likelihood (REML) in a series of bivariate animal model evaluations with three generations of pedigree in WOMBAT (Meyer 2007). For Angus ≈ 52% of the CIMF records had a corresponding HIMF record. However, no Angus bulls or Herefords with a scan IMF record had a corresponding CIMF record.

The model fitted for CIMF had fixed effects of CG, linear and quadratic effects of carcase weight as covariates and a random additive genetic effect of animal. CG were defined as per Graser *et al.* (2005). Models fitted for BIMF and HIMF included the random additive genetic effect of animal and sire x herd as a random effect. The model also included the fixed effects season of birth (2 levels, summer and winter), sex (fitted to HIMF, 2 levels), dam age (scaled to 5yrs old) x season, dam age squared x season, heifer factor deviation x season (if the dam was a heifer age was deviated from 2yrs old), and age (centred at 500 days) x sex

RESULTS AND DISCUSSION

Means and Variation. In Angus the mean CIMF was 8.32% compared to means of 3.28% and 4.86% for BIMF and HIMF, respectively (Table 1). The standard deviation for the scan IMF traits was also lower than observed for CIMF. Removing the lean CG from the scan records increased the mean to 3.84% and 5.18% for BIMF and HIMF, respectively without noticeably reducing standard deviation.

Hereford CIMF had a mean of 4.29% which was higher to the mean for BIMF (3.20%) and HIMF (3.83%; Table 2). However, the variation in CIMF (sd. of 2.16%) was noticeably greater than observed for BIMF (1.35%) and HIMF (1.65%; Table 2). As observed for Angus scan IMF traits, removing the lean CG increased the mean without significantly reducing the standard deviation (Table 2) for scanned traits.

Genetic Variation and Heritability. The heritability of CIMF in Angus was moderate (0.32) (Table 2) and similar to earlier estimates from Angus BREEDPLAN data (Reverter and Johnston 2001; Bȍrner *et al.* 2013). The heritability of CIMF in Herefords (0.37; Table 2) aligns with the observation in Angus and earlier estimates from the Hereford BREEDPLAN data (Reverter and Johnston 2001, Meyer *et al.* 2004).

Additive genetic variation for BIMF and HIMF in Angus was lower than observed for CIMF (Table 1). The all data BIMF and HIMF records for Angus had heritability estimates of 0.17 and 0.27, respectively and were similar to previous estimates of Reverter and Johnston (2001) and

Bõrner *et al.* (2013). Removing the leaner CG from the Angus BIMF records led to a slight increase in the additive genetic variance and heritability (0.21; Table 1). The heritability of HIMF was not improved by removing CG for a rib fat of 3mm and P8 fat of 5mm (Table 1). However, if fat depth thresholds were set at a rib fat of 5mm and P8 fat of 7mm the heritability of HIMF increased to 0.33 (Table 1).

The heritability of BIMF in Herefords was estimated at 0.20 (Table 2) which was slightly lower than previous estimates using Hereford BREEDPLAN data (Reverter and Johnston 2001, Meyer *et al.* 2004). Removing the lean CG increased the heritability of BIMF and HIMF to 0.23 and 0.30, respectively (Table 2). As observed for Angus heifers, using the more stringent fat depth cut offs resulted in a larger increase in the heritability estimate for HIMF (Table 2).

Table 1: Estimates of additive genetic variance (σ^2_a), heritability (h^2) of scan IMF traits and genetic correlation (r_g) between scan and carcase intramuscular fat (CIMF) for Angus

Subset	Records	% of data	Mean (%)	SD	σ^2_a	h^2	r_g CIMF
			Carcase IMF				
All data	7,833	100	8.32	3.90	1.324 ± 0.202	0.32 ± 0.05	
			Bull IMF (fat cut offs BRF=3,BP8F=5)				
All data	226,687	100	3.28	1.49	0.163 ± 0.007	0.17 ± 0.01	0.37 ± 0.11
Subset 1	120,636	53	3.84	1.38	0.190 ± 0.010	0.21 ± 0.01	0.48 ± 0.13
Subset 2	149,122	66	3.61	1.50	0.198 ± 0.010	0.19 ± 0.01	0.47 ± 0.12
			Heifers and steer IMF (fat cut offs HRF=3,HP8F=5)				
All data	245,840	100	4.86	1.85	0.395 ± 0.009	0.27 ± 0.01	0.36 ± 0.07
Subset 1	204,551	83	5.18	1.74	0.427 ± 0.010	0.29 ± 0.01	0.37 ± 0.07
Subset 2	235,580	96	4.93	1.84	0.408 ± 0.010	0.27 ± 0.01	0.36 ± 0.07
			Heifers and steer IMF (fat cut offs HRF=5,HP8F=7)				
Subset 1	137,850	56	5.60	1.62	0.444 ± 0.013	0.33 ± 0.01	0.39 ± 0.08
Subset 2	162,373	66	5.31	1.80	0.440 ± 0.012	0.28 ± 0.01	0.37 ± 0.08

Genetic Correlations. If all BIMF records were incorporated in the bivariate analysis of BIMF and CIMF for Angus, the genetic correlation estimate was 0.37 (Table 1). This estimate was higher than correlations reported by Reverter and Johnston (0.13; 2001) but similar to the estimates by Bõrner *et al.* (2013) in young (0.34) and older bulls (0.43). By removing lean CG, the genetic correlation estimate between BIMF and CIMF in Angus increased to 0.48 (Table 1). Reintroducing some of the lean CG that had IMF variation in the top 25% lead to only a small reduction in the genetic correlation compared to when all lean CG are removed (0.47) and remained noticeably higher than when all data was included (Table 1). Increasing the variation threshold to include CG in the top 50% resulted in a genetic correlation of 0.39 (not presented). The results suggest that applying a threshold based on minimum fat depth to scan IMF data resulted in higher estimates of the genetic associations between scan and carcase IMF.

Removing lean CG from the Angus HIMF records did not improve the genetic correlation with CIMF (Table 1). While using more stringent fat cut offs (rib fat < 5mm P8 fat < 7mm) leads to increases in the additive variance and heritability of HIMF the increase in the genetic correlation with CIMF was minimal (Table 1). The correlation between HIMF and CIMF in the Angus industry data was previously reported at 0.45 (Reverter and Johnston 2001) and 0.39 (mean age 443 days) and 0.42 (583 days; Bõrner *et al.* 2013).

In Herefords, removing data selectively did not result in increases in either heritability estimates for BIMF or HIMF, or in the genetic correlations between BIMF and CIMF, or HIMF and CIMF (Table 2). The genetic correlation between BIMF and CIMF in Hereford was 0.69 (all data) which was slightly stronger than the previous estimate of 0.59 presented by Meyer *et al.*

(2004). This may, in part, be due to Hereford bulls being scanned on average 30 days older and 0.5mm fatter over the rib then the Angus Bulls. The genetic correlation between HIMF and CIMF in Herefords was considerably lower than the previous estimate presented by Meyer *et al.* (2004) of 0.97.

The estimates of the genetic correlation between scan IMF and CIMF by Bȍrner *et al.* (2013) and within this study suggest that scanning younger and leaner cattle will reduce the strength of the association. This may, in part, be due to a decline in the accuracy of the scan equipment when measuring lean cattle, but further research will be required to test this hypothesis.

Table 2: Estimates of additive genetic variance (σ^2_a), heritability (h^2) of scan IMF traits and genetic correlation (r_g) between scan and carcase intramuscular fat (CIMF) for Hereford

Cut off	Records	% of data	Mean (%)	SD	σ^2_a	h^2	r_g CIMF
			Carcase IMF				
All data	1,836	100	4.29	2.16	0.46 ± 0.14	0.37 ± 0.10	
			Bull IMF (fat cut offs BRF=3,BP8F=5)				
All data	86,603	100	2.93	1.35	0.12 ± 0.01	0.20 ± 0.01	0.69 ± 0.17
Subset 1	63,274	73	3.20	1.28	0.14 ± 0.01	0.23 ± 0.01	0.61 ± 0.18
Subset 2	69,931	81	3.10	1.33	0.14 ± 0.01	0.22 ± 0.01	0.62 ± 0.18
			Heifers and steer IMF (fat cut offs HRF=3,HP8F=5)				
All data	70,020	100	3.83	1.65	0.30 ± 0.01	0.28 ± 0.01	0.54 ± 0.16
Subset 1	62,208	89	4.02	1.59	0.32 ± 0.01	0.30 ± 0.01	0.48 ± 0.15
Subset 2	64,250	92	3.97	1.62	0.32 ± 0.01	0.30 ± 0.01	0.47 ± 0.16
			Heifers and steer IMF (fat cut offs HRF=5,HP8F=7)				
Subset 1	44,657	64	4.26	1.53	0.36 ± 0.02	0.33 ± 0.02	0.55 ± 0.18
Subset 2	51,049	73	4.06	1.64	0.36 ± 0.02	0.30 ± 0.01	0.46 ± 0.18

CONCLUSION

The continuing trend towards scanning Angus bulls at 400-days and at leaner subcutaneous fat depths is causing a decline in the genetic correlation between scan and carcase IMF. Removing contemporary groups based on fat depth thresholds resulted in a slight strengthening of the genetic correlation between scan and carcase IMF in Angus bulls. Producers should avoid scanning herds with fat levels below the cut offs presented, therefore allowing animals the opportunity to express their genetic merit for IMF. Increasing the number of CIMF records is desirable, yet difficulties in obtaining abattoir progeny test data mean there is also a need to improve the quality of scan IMF records. Alternatively the genetic correlations between scan and carcase IMF in BREEDPLAN evaluations should be adjusted, which will reduce the utility of scanning, but due to the large number of animals that can be scanned and the relative low cost of measuring scan IMF it still remains the most practical correlated trait for CIMF.

REFERENCES

Bȍrner, V., Johnston, D.J and Graser, H. U., (2013) *Ani. Prod. Sci.* **53**:1075

Graser, H-U., Tier, B., Johnston, D.J. and Barwick, S.A. (2005) *Aust. J. Exp. Agric.* **45**:916

Johnston, D. J., Tier, B., Graser, H. U., and Girard, C. (1999) *Proc. Assoc. Advmt. Anim. Genet.,* **13:**193

Meyer, K., Johnston, D.J and Graser, H. U., (2004) *Aus. J. Ag. Res.* **55**:195

Meyer, K. (2007). *J. Zhejiang Uni. Sci.B*, ***8***:815

Reverter A., Johnston, D. J., Graser, H-U., Wolcott, M.L. and Upton W.H. (2000) *Anim. Sci.* **78**:1786

Reverter A. and Johnston, D. J. (2001) *Proc. Assoc. Advmt. Anim. Genet.,* **14:**159

Proc. Assoc. Advmt. Breed. Genet. 21: 257-260

INTEGRATION OF GENOMIC INFORMATION INTO NATIONAL CATTLE AND SHEEP EVALUATIONS – PAST, PRESENT AND FUTURE

D.P. Berry[1], J.F. Kearney[2], N. McHugh[1], E. Wall[3], D. Purfield[1], and A.R. Cromie[2]

[1] Teagasc, Moorepark, Ireland, [2] Irish Cattle Breeding Federation, Ireland, [3]Sheep Ireland, Ireland

SUMMARY

National (two-step) genomic evaluations for Holstein-Friesian dairy cattle have been available in Ireland since 2009; retrospective analysis reveals minimal bias in prediction and more accurate rankings compared to pedigree index. The size of the training population for dairying (up to 5,105 animals) has increased 5-fold in recent years through sharing of genotypes. One-step multi-breed genomic predictions for beef cattle will become official in autumn 2015 made possible through a genotyped population of over 100,000 individuals. Genotypes of all cattle are from the custom International Dairy and Beef genotype panel developed in Ireland; this panel consists of 18,217 SNPs. Genotyping of 11,400 sheep for the development of genomic predictions has just begun.

GENOMIC EVALUATIONS IN DAIRY

Past. Genomic selection was launched in Ireland in February 2009 for animals that were >87.5% Holstein-Friesian, but no more than 12.5% Friesian; the cost was €50 per animal and was freely available to all (i.e., both national and international farmers, breeders, breeding companies). Analyses revealed poor accuracy of genomic predictions in pure Friesians at the time so an initiative was undertaken to genotype high reliability pure Friesian males; genomic predictions for Friesians were available in 2010. The initial training population consisted of 998 high reliability Holstein-Friesian AI sires. Predictions were (and still are) undertaken based on a two-step approach using custom developed software. Predicted transmitting abilities (PTAs) from domestic genetic evaluations, weighted by their reliability were the dependent variables; later years used MACE evaluations as the dependent variable. The number of animals in the training population increased in the following years through bilateral sharing agreements and in 2015 includes up to 5,105 animals. Sharing was (and continues to be) on the basis that genotypes will not be passed onto third parties, but more importantly that no restrictions are imposed on who can obtain a genomic evaluation. For example, genomic predictions are undertaken by the Irish Cattle Breeding Federation (ICBF) for thousands of animals annually for non-indigenous breeding companies.

Present. Once a bull has >50% reliability for direct calving difficulty (a crucial trait in Ireland) in Ireland, then that bull is available for widespread use. Ireland maintains a national breeding program (i.e., G€N€ IR€LAND) to ensure robust genetic evaluations especially for calving difficulty and monitoring of congenital defects. No financial incentives exist for G€N€ IR€LAND farmers other than reduced semen cost. The most recent daughter-based national genetic evaluation of 182 bulls genomically tested between the years 2009 and 2012 but now with >70% reliability for milk production and reproduction are in Table 1; the correlation with their respective pedigree index predictions and their first genomic predictions in their year of sale are in Table 1.

Genotypes of all cattle are generated using the custom International Dairy and Beef (IDB) genotype panel developed in Ireland. This panel includes 18,217 SNPs which include an extra 5,765 SNPs, additional to the Illumina low-density base content for imputation to higher density (especially in beef). Also included are 1,927 SNPs for imputation to microsatellites for parentage testing as well as four lethal mutations and causal (or tightly linked) SNPs in major genes associated with performance or congenital defects. A total of 2,973 SNPs as part of on-going research projects are also included on the panel.

Table 1. Mean (reliability) daughter-based PTAs from the most recent genetic evaluation as well as past parental average (PA) or genomic-based PTAs; also included is the correlation with the most recent daughter PTA for both PA and genomic PTAs

	Mean			Correlations	
Trait	Daughter	PA	Genomic	PA	Genomic
Milk (kg)	116 (90)	168 (41)	108 (61)	0.71	0.79
Fat (kg)	10.4	11.9	10.2	0.55	0.7
Protein (kg)	7.8	9.6	7.7	0.63	0.75
Fertility (d)	-4.5 (71)	-3.1 (30)	-3.7 (46)	0.6	0.63
Survival (%)	2.01	1.52	1.7	0.41	0.63

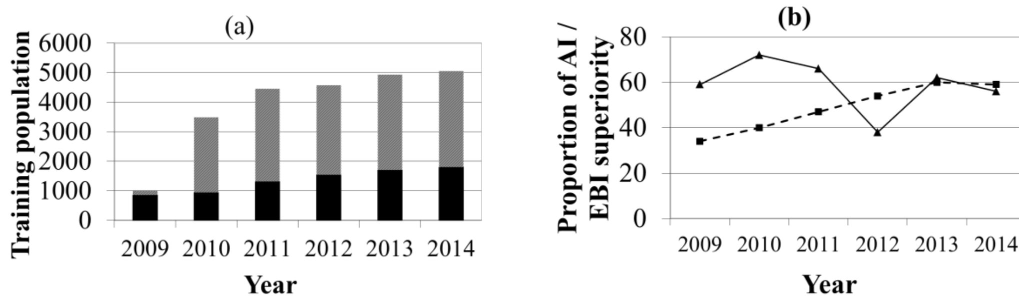


Figure 1. (a) Number of animals in the training population generated within Ireland (Black) or from bilateral sharing (grey) and (b) proportion of semen sales from genomically tested sires (broken line) and genetic superiority (€) of used genomically tested sires relative to used proven bulls (continuous line); standard deviation of national index (EBI) is €72.

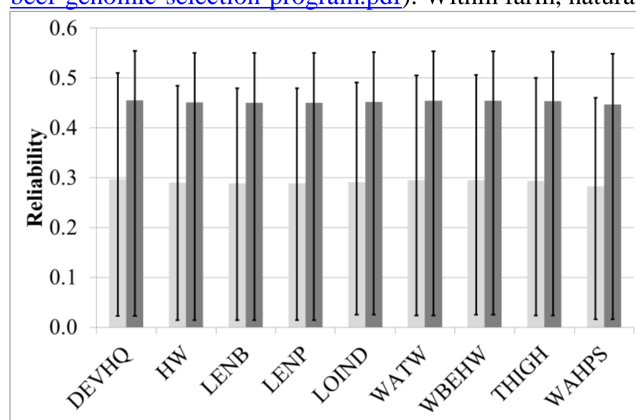
Future. Attempts to exchange Jersey genotypes were largely unsuccessful; therefore an initiative began in 2014 to genotype commercial crossbred Jersey cows with the view to developing genomic predictions for Jersey animals in 2015. Farmers paid half the cost of the genotyping; 2,811 genotypes from Jersey crossbred cows now exist with more being genotyped in 2015. Research is currently underway on the potential of undertaking a one-step multi-breed genomic evaluation for dairying with possible implementation in spring 2016. Ireland is part of the 1000 Bull Genomes Project and has imputed all dairy genotypes to full sequence. Genome-wide association studies are underway to attempt to more closely track the causal mutations; once useful information is detected it will be added to updated versions of the IDB. Research is also underway on generating estimates of genetic/genomic merit for difficult to measure traits. One such initiative is the global dry matter initiative (gDMI) where across country genomic predictions for feed intake have been undertaken (de Haas et al., 2015) and SNP effects on each country's scale were generated. Results showed a benefit from sharing of phenotypic and genomic information, especially in populations where the phenotypic information was lacking. Interest in Ireland is intensifying on the potential of precision genomic matings especially as part of the national sire advice algorithms.

GENOMIC EVALUATIONS IN BEEF

Past. The initial analysis in 2013 with 4,233 high density genotypes from multiple beef breeds revealed an insufficient population size to derive meaningful genomic predictions. In some instances a negative correlation existed between genomic predictions and progeny-based predictions when the breed in the validation dataset was not included in the reference dataset. A considerably large dataset of phenotyped and genotyped cattle was necessary to implement an accurate multi-breed genomic selection program. With an expectation of changing parentage

testing in cattle from microsatellites to SNPs, breed societies were requested to genotype all male calves using the IDB panel in 2014. Considerable transfer of germplasm exists between dairy and beef herds in Ireland.

Present. Funding from the EU to Irish farmers has, for the last 5 years, been conditional on the recording of sire of the calf born and additional phenotypic information; it is a legal requirement to record the dam. In 2014, farmers were also required to genotype at least 15% of their cow herd; animals were optimally selected from the national database to aid in the derivation of genomic predictions – selection was based on genomic diversity, phenotypic diversity and information content (http://www.icbf.com/wp/wp-content/uploads/2014/05/Selection-of-animals-for-use-in-beef-genomic-selection-program.pdf). Within farm, natural mating sires were prioritised.



Following edits, 105,561 genotypes imputed to higher density, were available on AI, pedigree and commercial animals. The reliability of genetic evaluations for a series of linear type traits increased when a one-step genomic evaluation was undertaken (Figure 2). Within breed correlations between the genomic EBVs and progeny based EBVs varied from 0.51 to 0.73. No difference existed in predictive ability between high density or 50,000 SNPs.

Figure 2. Mean reliability of young bulls based on pedigree index (light bar) and genomic predictions (dark bar) for a selection of type traits; min and max represented as error bars.

Future. Focus is on difficult to measure traits (e.g., feed intake and meat quality); for example it is hoped to genotype and phenotype over 15,000 animals for meat quality in the next 24 months. Because of the short shelf-life of minced-meat, amplicon sequencing with several hundred SNPs for the determination of Angus breed proportion is under investigation to quicken turn-around time for genotyping. The Irish dairy herd is expected to expand by approximately 30%; as reproductive performance improves with selection on the national dairy breeding objective, the scope for mating dairy cows with beef semen is greater. Hence, a breeding program for beef bulls to use on dairy cows has begun.

GENOMIC EVALUATIONS IN SHEEP

Present. Funding to genotype 11,400 sheep exists. In total, 363 highly unrelated sheep from 18 different "breeds" in Ireland have been genotyped on the Illumina high density ovine beadchip. This is to establish breed differentiation and quantify the potential of across breed genomic predictions. Following discussions on whether to genotype commercial crossbred animals with good phenotypes or purebred pedigree animals, it was decided to genotype pedigree animals at the top of the breeding pyramid to, amongst other reasons, maximise the likelihood of genotyping continuing beyond the lifetime of the project. Ear biopsies of all pedigree animals (i.e., rams, ewes and lambs) are currently being collected from participating farms who record information in the national database. Influential parents will be genotyped using the Illumina high density beadchip, less influential parents will be genotyped using the Ovine50 beadchip and the remaining parents will be genotyped using a lower density panel being developed by an international consortium.

Future. A multi-breed genomic evaluation will be attempted. The potential usefulness of an ultra-low density, ultra-low cost genotype panel for parentage assignment (in commercial or progeny test animals) and possible screening for some structural mutations of interest (e.g., scrapie) will also be investigated. Sharing of Texel genotypes with the UK has begun. Suffolk EBVs from both the UK and Ireland have also been exchanged. Table 2 summarises the correlations between EBVs and inferred genetic correlations (following adjustment for accuracy of the respective national EBVs; Calo et al., 1973). The strong correlations that exist between some traits suggest a possible advantage in augmenting the accuracy of genomic evaluations by using the (weighted) UK EBVs as correlated traits in the Irish genomic prediction algorithms. No investment has yet been made in sequencing of sheep in Ireland.

Table 2. Number of animals (n), correlation between EBVs (r_{EBV}) and inferred genetic correlations (r_g) for a series of traits between the UK and Ireland

Trait	n	r_{EBV}	r_g
8-wk weight	222	0.254	0.937
Weight at scanning	222	0.190	0.648
Ewe weight	222	0.012	0.073
Scanned muscle depth	210	0.075	0.258
Scanned muscle fat	210	0.175	0.666
Number of lambs born	215	0.450	0.993

LESSONS LEARNT AND FUTURE PLANS

Not having a national repository for biological samples or DNA bank was a short-sight. Since 2010, two semen samples have to be provided, free-of-charge, to the ICBF for all AI bulls used in Ireland. The decision to develop a custom genotype platform was the correct decision; it did not require considerable effort (approximately 12 weeks to develop although validation of genotypes for some mutations is on-going), or alignment to any one genotyping platform or service provider; it did provide greater flexibility on what variants to include, and the cost per genotype was lower than could be provided by any other service provider with an available custom genotype platform. The list of SNPs included on the international dairy and beef chip is available from the author.

The development of genomic selection in the Irish dairy population cost approximately AUS$0.55 m (including research personnel but excluding the required IT developments for reporting); this was predominantly funded by competitive research grants and bilateral sharing of genotypes. Genomic selection has delivered AUS$23m to the dairy sector since 2009. Genomic selection in sheep (excluding the cost of phenotyping and IT developments) will cost approximately AUS$0.96 m almost exclusively from a competitive research grant. Genomic selection in beef has cost €5.7 m, acquired from competitive research grants (€0.75 m) or EU farmer support mechanisms. Although it does require public investment, retaining public ownership of genomic proofs and the free access to genomic predictions is a considerable advantage. Sharing of genotypes with minimal conditions is also key to success.

Like mobile phones and computers, it appears that the cost of genotyping using arrays is unlikely to reduce considerably in the near future but instead the quantity of SNPs to be included will increase. This will avoid the necessity to impute, improving computing time and accuracy.

REFERENCES

Calo, L.L., McDowell R.E., VanVleck L.D. and Miller P.D. (1973) *J. Anim. Sci.* **37**:676

de Haas, Y, Pryce JE, Calus MPL, Wall E, Berry D.P., Løvendahl P., Krattenmacher, N, Miglior, F, Weigel, K, Spurlock D, Macdonald K, Hulsegge B, and Veerkamp, RF (2015) *J. Dairy Sci.* (In press)

Proc. Assoc. Advmt. Breed. Genet. 21: 261-265

SOLE HAEMORRHAGE, DERMATITIS AND Co. - HOW GENOMIC INFORMATION AND PRECISE PHENOTYPES HELP TO UNSCRAMBLE GENETIC BACKGROUND OF HEALTH TRAITS IN DAIRY CATTLE

K. Schöpke

Institute of Agricultural and Nutritional Sciences, University of Halle, Halle, Germany
Moritzburg State Stud, Saxon State Stud Administration, Moritzburg, Germany

SUMMARY

Disorders of the bovine hoof are important factors influencing the well-being and milk production of the dairy cow. Here, results from 2 different studies were used to demonstrate the value of how contemporary groups are defined, standardized recording, and improved trait definitions, to investigate the genetic background of important claw disorders. In the first study, 1,962 first-lactation cows from 7 commercial (contract) herds were subjected to hoof trimming with an assessment of hoof disorders as binary traits. Sole hemorrhage (SH), white line disease, sole ulcer, and interdigital hyperplasia (IH) showed to be the most important noninfectious claw disorders. The DNA of 1,183 of the cows was used for analyses with a custom-made array of 384 SNP. It revealed that SNP rs29017173 is significantly associated with SH disorder status. For IH, bull lineages with high proportions of IH affected daughters and granddaughters were identified. With the help of 192 genotyped cows (Illumina BovineSNP50 BeadChip) from well-selected cohorts, 4 candidate regions were identified. A second study based on 729 pregnant heifers from an US-American commercial dairy herd focused on bovine digital dermatitis (DD). New DD trait definitions were used to investigate the genetic background of this infectious disease. The new traits enabled the differentiation of clinical stages and their succession over time. The heritability estimates for the DD traits ranged between 0.19 and 0.52. An association study, based on 106 genotyped cows from this study, revealed 3 promising candidate regions.

INTRODUCTION

The implementation of genomic selection by large parts of the animal breeding industry has enabled gains in predicting the accuracy of breeding values, increased genetic progress, and allowed new breeding strategies. With more precise knowledge of the genetic information, it seems that the limitation is now phenotype availability. This limit is caused by either the quality or the quantity of phenotypes. In general, precise phenotypes and standardized data management are key aspects for high quality data. New or hard to measure traits often result in a small, or hardly representative sample. Health traits have recently been growing in importance, but are often especially challenging (Egger-Danner *et al.* 2015). Most farms record disease events as they occur and thus the successful principal to work with defined contemporary groups at a specific point in time is generally neglected.

This paper presents the use of defined contemporary groups, standardized recording, and improved trait definitions, to investigate the genetic background of and to identify associated genetic regions for the claw disorders sole hemorrhage, interdigital hyperplasia, and bovine digital dermatitis in dairy cattle.

MATERIALS AND METHODS

This paper is based on two different trials.

Trial_1: Seven commercial herds in north-eastern Germany were selected from a pool of contract dairy herds of the breed association Mecklenburg-Vorpommern. All herds had similar housing conditions and feeding. First lactation cows being in a similar stage of lactation were defined as contemporary groups. A fixed team of 3 people assessed a total of 1,962 first-lactation Holstein cows during 24 herd visits. Claw diagnoses were recorded for hind legs at the time of hoof trimming. The disorder status included assessing clinical and subclinical claw disorder cases as binary traits. A detailed description of the trial was described in Schöpke *et al.* (2013).

About half of the cows (1,183) were genotyped using a custom-made 384 array.

Before this study, no published QTL for sole hemorrhage (SH) were available. Thus, the selection of SNP for the custom-made array followed the general assumption that genetic correlation between leg conformation traits and claw disorders exists.

SNP selection was based on a 4-step-strategy that included an in silico analysis of published QTL associated with conformation traits in cattle and pigs; the identification of syntenic chromosomal regions across cattle, pigs, and human; the selection of candidate genes; and the selection and validation of SNP. The list of candidate genes contained 1,035 genes that were assigned to at least one of the biological functions development, function and disease of skeleton, muscles, and connective tissue; cell signalling; disorder in vitamin and mineral metabolism; or hydrate metabolism. For the array, SNP within 384 of these genes were selected according the following criteria: validation for NCBI or bovine 50K chip, SNP interval (0.7–1.0 Mb), minor allele frequency (>0.05), not located in a repeat region, and tested on a Holstein-Frisian population. With regard to the selection of cows for genotyping with the custom-made array, entire contemporary groups were selected that exhibit a “normal” range of prevalence rates for the claw disorder sole hemorrhage. All further samples were genotyped for 1 SNP that was strongly associated with SH by a fluorescence resonance energy transfer assay (Förster 1946). Detailed description for DNA extraction, SNP selection, and genotyping can be found in Swalve *et al.* (2014). In a second part of this trial, contemporary groups with the highest prevalence rates for interdigital hyperplasia (IH) were used as a basis for genotyping 192 cows with the Illumina BovineSNP50 BeadChip. as described in Sammler *et al*. (2015).

Trial_2: In this trial, 729 pregnant (nulliparous) heifers from a commercial Holstein dairy herd in Wisconsin, USA that was endemically affected by bovine digital dermatitis (DD), were inspected in a stand-up chute on a regular basis: at least 3 times per heifer within a mean (SD) individual cow observation time of 176 days (20.1). In total 6,444 clinical observations for DD were collected applying the M-score system as defined by Döpfer *et al.* (1997) and Gomez *et al.* (2014). This system is a classification scheme for stages of DD that allows a macroscopic scoring based on clinical inspections of the bovine foot, thus it describes the stages of lesion development. Briefly, lesions were classified as M0 for unaffected animals with no clinical lesions; as M1 for infected heifers with early lesions smaller than 2 cm in diameter (non-active); and as M2 for infected heifers with a classic active lesion of >2 cm of diameter considered to be infectious. An M4 stage denotes late and chronic stages of DD with (M4.1) or without (M4) small (<2 cm diameter) M1 lesions within their perimeter. M-scores were used to define new DD trait definitions with different complexity (Table 1) as described in Schöpke *et al.* (2015). Trait TBIN denoted a very basic description of the clinical DD status. This binary trait separated between unaffected heifers throughout the entire observation period and all other heifers. A special consideration of heifers reaching an active stage of lesion (M2 or M4.1) was given with trait TBINA. Trait TSEVCAT described the severity of DD cases a heifer was afflicted by in two slightly different versions. TSEVCAT was a categorical trait with 3 classes comparing not affected heifers (always M0; score = 1) with heifers having at least one M1, M4, or M4.1 but never M2 (score=2), and cows suffering at least once from classic active ulcers (M2; score=3). TSEVCAT41

was very similar to TSEVCAT but differed concerning heifers with at least one M4.1 event, which received a score of 3. The known difference between M2-cow types (Gomez *et al.* 2014) is basis for the definition of trait TCTM2SC that classified heifers according the number of active M2 lesions during the observation period (score of 1=type I heifer: no M2 lesions; score of 2=type II heifer: exactly one M2 lesion, score of 3=type III heifer: multiple M2 lesions). TCTM2SC counted all M2 events considering every leg separately and TCTM2 counted per event date. Trait TTRANS accounted for the changes of the M-stages over successive evaluations and thus included a better description of the DD infection dynamics. For TTRANS, the transitions between stages were classified (1: staying not-affected; 2: healing, improving; 3: staying affected on the same/comparable stage; 4: aggravating), the classes were weighted, and a transition score was derived. To compute TTRANS, for each heifer all observations were classified for heifer type as explained above, weighted, summarized, and divided by the number of transitions observed. A reference scenario (TREF) was defined by considering the first evaluation of each heifer as the only information. TREF denotes a single scoring for DD as has been commonly used in most studies applying genetic-statistical methods.

For genotyping, the Illumina BovineSNP50 BeadChip was used to genotype 63 animals; another 43 animals were genotyped with the BovineHD Genotyping Bead Chip (777K).

Table 1. Description of trait definitions for digital dermatitis, number of observations in the final data set, means, and standard deviations (SD); category frequencies instead of means for traits with more than two categories

Trait	Trait definition	No. of observations (no. of cows)	Mean	SD
TBIN	Binary trait that differentiates between consistently not DD affected cows (0) and cows with at least one observation with a DD lesion (1)	729 (729)	0.52	0.50
TBINA	Binary trait that differentiates between cows that never (0) / at least once (1) experience an active stage of DD lesion	729 (729)	0.40	0.49
TSEVCAT	Categorial trait that differentiates between three severity categories of DD lesions: consistently not affected (1), at least once M1, M4, or M4.1 (2), at least once M2 (3)	729 (729)	(1) 48.3% (2) 15.7% (3) 36.0%	
TSEVCAT41	Categorial trait that differentiates between three severity categories of DD lesions: consistently not affected (1), at least once M1 or M4 (2), at least once active stage M2 or M4.1 (3)	729 (729)	(1) 48.3% (2) 11.9% (3) 39.8%	
TCTM2	Categorial trait that differentiates between three DD cow types concerning the number of M2 events: never M2 (1), once M2, at least twice M2 (3)	691 (691)	(1) 64.0% (2) 14.5% (3) 21.5%	0.79
TCTM2SC	Categorial trait that differentiates between three DD cow types concerning the number of M2 events considering legs separately; never M2 (1), once M2, at least twice M2 (3)	691 (691)	(1) 64.0% (2) 18.4% (3) 17.6%	0.83
TTRANS	Transition score for the classified and weighted transitions between DD stages	729 (729)	17.90	9.05
TREF	Binary trait that differentiates the first observation of the cow into not affected (0) or affected (1)	729 (729)	0.12	0.33

Statistical Analyses: Data preparation, editing, and examination of alternative modelling of fixed effects as well as preliminary χ^2 tests for genotypic associations with SH, IH, or DD were conducted using the statistical package SAS 9.1 and 9.4. Variance components were estimated using a restricted maximum likelihood (REML) animal model and applying the ASReml 3.0 software package. Associations between genotype and disorder status were also tested using PLINK software.

RESULTS AND DISCUSSION

Trial_1: The 4 most important noninfectious disorders were sole hemorrhage, white line disease, sole ulcer, and interdigital hyperplasia with SH being the predominant disorder. Prevalence rate varied between 5.5 (IH) and 57.3 % (SH) showing remarkable differences of within-herd levels (Table 2). The prevalence level of SH is higher than in other studies, which is largely because "mild" or subclinical cases of diseases were included. For SH, herd-visit date, stage of lactation, and body weight significantly affected the probability of occurrence and thus were included as fixed effects in the model when accounting for SNP genotype effects. Analyses by PLINK of 295 SNP (MAF>0.5) revealed a highly significant association (P<0.001) between disorder status for SH and the SNP (HAPMap54883-rs29017173) within the IQGAP1 gene (BTA 21). The GLIMMIX analyses resulted in back-transformed means of the disorder status of 0.35 (AA), 0.49 (AG), and 0.54 (GG) when comparing the 3 genotypes in a reduced data set (herd-visit cohorts with extreme frequencies for SH were excluded). Using the full data set, the back transformed means of the SH status were 0.50 (AA), 0.56 (AG), and 0.60 (GG). Polymorphism of the SNP showed substantial effects for the occurrence of SH, but it was also found to be associated with feet and leg traits from the classical conformation score system. Fortunately, the same allele is favoured for all traits with substantial effects. IQGAP1 is proven via knock-out mice to play a critical role in postischemic neovascularization and tissue repair (Urao *et al.* 2010). Thus, it can serve as a promising candidate gene for the pathogenesis of SH in cattle. However, it might be more a question of tolerance here than of resistance.

When assessing the disorder status of interdigital hyperplasia (IH), 107 IH positive animals were identified. When separating the IH phenotype into one-side and pairwise affected rear legs, 71 cows and 36 cows were found, respectively. An investigation of sires with at least 5 daughters in the data set revealed an IH predisposition of 5 sires that are sons of the same bull. When comparing bulls with at least 140 granddaughters in the data set, the identified bull had a remarkably high proportion (9.4 %) of IH affected granddaughters. This result confirms the occurrence of IH in some bull lineages, this observation has occasionally been mentioned in literature (Hogreve 1964).

Table 2. Prevalence rates for the 4 most important noninfectious disorders and herd-visit prevalence interval

	All observations	Within herd - date of visit	
Name of disorder	[%]	Min [%]	Max [%]
Laminitis	57.3	25	92
White line disease	12.6	2	32
Sole ulcer	7.1	0	25
Interdigital hyperplasia	5.5	0	20

Within the 192 genotyped cows, 87 were IH positive (56 one-side affected vs. 31 pairwise affected) and 105 were negative (=controls). A case-control study on the genetic background of IH revealed associated regions on 4 different chromosomes.

Trial_2: From all 6,444 observations, 68.0 % of the records were found to be negative, i.e. "healthy" regarding DD. Out of the 32 % DD-positive observations 54.8 % showed a chronic stage (M4 or M4.1) of which 37.7 % were chronic and active (M4.1). The infectious stages M2 and M4.1 together accounted for 11.3 % of all observations. 48.2 % of the heifers were consistently not affected by DD during the entire observation time. Estimates for heritabilities from univariate models were 0.19±0.11 (TBIN), 0.20±0.11 (TBINA), 0.27±0.12 (TSEVCAT), 0.23±0.12 (TSEVCAT41), 0.46±0.16 (TCTM2SC), 0.52±0.17 (TCTM2), 0.42±0.15 (TTRANS), and 0.19±0.11 (TREF). Estimates of heritabilities for DD exist in the literature (e.g. van der Linde *et al*. 2010; Gernand and König 2014) however, none of the published studies used M-scale scored records. Estimates in the present study were higher than results from literature. This might be due to the limited sample size. Even though, the results presented are based on a well-established classification system for clinical stages and their succession over time, and thus serve as a comparison between trait definitions that reach beyond a separation of affected and non-affected animals. Those traits were also basis for the association study. Analyses by PLINK showed candidate regions on 3 different chromosomes. For 1 of the regions, a haplotype consisting of several SNP being in complete linkage disequilibrium was identified.

Conclusion. Phenotyping for health traits appears to be very difficult. This is partly due to dominant environmental effects and the occurrence of genotype by environmental interaction, but it is caused also by the difficulties of precisely defining "diseased" or "healthy" states. In Trial_1, an association between SH and the IQGAP1 gene was identified. The strength of this association differed between the full and the reduced data set, and thus it demonstrated that the detection of genetic effects or individual gene effects is strongly dependent on a well-planned study design, implying selection of herds, definition of phenotype, and time of evaluation. Trial_2 demonstrated the application of the M-scale scoring system when analysing the genetic pre-disposition for DD. This system is widespread in the veterinary field. Its first use for genetic analyses revealed higher heritability estimates than known from previous studies, and therefore the genetic predisposition for DD might be higher than previously assumed. The findings indicated that applying this improved phenotype definition of DD allows for improved management strategies and improved strategies for genetic selection.

REFERENCES

Döpfer, D., Koopmans A., Meijer F.A., Szakáll I., *et al*. (1997) *Vet. Rec.* **140**:620.
Egger-Danner C., Cole J.B., Pryce J.E., Gengler N., *et al*. (2015) *Animal* **9**: 191.
Förster, T. (1946) *Naturwissenschaften* **33**:166.
Gernand E. and König S. (2014) *J. Dairy Sci.* **97**:5872.
Gomez A., Anklam K.S., Cook N.B., Riemann J., *et al*. (2014) *J. Dairy Sci.* **97**:4864.
Hogreve F. (1964) *Tierärztliche Umschau* **9**: 453.
Sammler S., Scheper C., Wensch-Dorendorf M., Schöpke K., Brenig B. (2015) *in prep.*
Schöpke K., Gomez A., Dunbar K.A., Swalve H.H., and Döpfer D. (2015) *J. D*airy Sci. accepted.
Schöpke, K., Weidling S., Pijl R., and Swalve H.H. (2013) *J. Dairy Sci.* **96**: 679.
Swalve H.H., Floren C., Wensch-Dorendorf M., Schöpke K., *et al*. (2014) *J. Dairy Sci.* **97**: 507.
Urao N., Razvi M., Oshikawa J., McKinney R.D., *et al*. (2010) *PLoS ONE* **15**: 5.
van der Linde C., de Jong G., Koenen E.P.C., and Eding H. (2010) *J. Dairy Sci.* **93**: 4883.

THE OPPORTUNITIES AND CHALLENGES OF INTEGRATING GENOMICS IN A BROILER BREEDING PROGRAM

R.J. Hawken, R.L. Sapp, R. Okimoto, J. Chen, R. Borg, C-H. Huang, T. Wing, C. Wang, J. Henshall, and M. Abrahamsen
Cobb-Vantress, Siloam Springs, Arkansas, 72761, USA

SUMMARY

Cobb-Vantress is one of the leading global suppliers of broiler breeding stock, with products distributed in more than 100 countries. Cobb has continually invested in new technologies to consistently deliver genetic improvement that provide a competitive advantage in the market place. Recently, Cobb has made significant investments to implement a genomic selection program to complement the traditional breeding program. In addition to genomic selection, opportunities such as causative mutation detection, parentage testing and simple trait selection have been successfully implemented in various breeding programs within Cobb. There are many challenges involved in implementing these genomic technologies, including a simple but complex effort toward the logistics of sample collection and management from multiple pure-line populations at different geographical locations. The current state of the genome sequence presents some barriers to the successful use of these technologies in some instances; however there is currently some significant effort toward its improvement. We believe that genomic technologies are beneficial technologies to improve the genetics of our broilers.

INTRODUCTION

Cobb-Vantress, Inc. (Cobb) is the world's oldest broiler breeding company. Since 1916 Cobb-Vantress has contributed to the dynamic growth of the global poultry industry that has transformed chicken into a popular, affordable and healthy protein choice. Cobb maintains a pedigree program, ensuring continual genetic progress for a production pipeline where it creates multiple parent stock targeted toward the production of multiple products with different performance profiles, ranging from highly efficient, to high yielding, and slow growing broilers. These products are successfully produced in very diverse environmental, management and regulatory production systems globally. There are several challenges facing the poultry industry requiring the production of alternative broiler solutions for future markets. Some of these challenges include;

- Volatile global grain prices emphasizing the need for continual improvement in feed conversion and use of alternative feed products.
- Emerging market opportunities emphasizing the need for diversified products for new environments.
- Welfare and customer requirements driving the need for innovative products such as antibiotic free chicken.
- Governmental and regulatory changes requiring the need for unique breeds (such as slow growing lines) or management practices.

Genomics is a technology being investigated to help Cobb create broiler solutions to tackle some of these industry challenges. To date, Cobb has successfully integrated genomic technologies such as genomic selection, parentage testing, identification and elimination of deleterious alleles, and single gene tests. This paper will address some of the challenges and opportunities that Cobb has identified through its genomic program.

CHICKEN BREEDING AT COBB

Over 2 million pure line chicks are hatched at Cobb annually. All chicks hatched on one of our seven pedigree farms are individually identified, and individually phenotyped for over 50 traits, including;

- broiler traits such as weight, feed conversion and breast meat percentage,
- reproduction traits such as hatch of fertile and egg production
- welfare and health traits such as skeletal defects, foot pad dermatitis and liveability

Less than 5% of hatched chicks are retained as future breeders based on selections using both phenotype and BLUP breeding values. It is estimated that one selected female pedigree breeder makes a genetic contribution over 3 million broilers (which are a four-way line cross). Given our ability to pedigree millions of birds and maintain high selection intensities we can make genetic gains very quickly. Figure 1 illustrates the progress made in the final broiler product (a cross of 4 pure-lines) over a 20 year period.

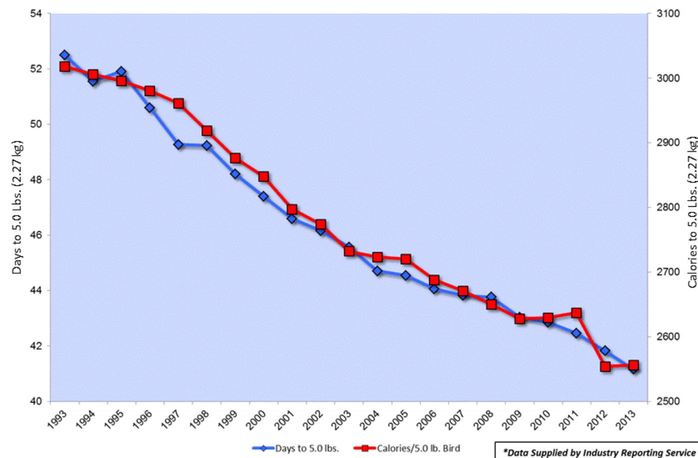


Figure 1. Days to 5.0 Lbs. (2.27 kg) and Calories/5.0Lb for broiler between 1993 and 2013

GENOMIC TOOLS FOR CHICKEN BREEDING

The chicken genome sequence was made available by the international chicken genome consortium in 2004 (Hillier *et al*. 2004) and has been revised three times (2006, 2011 and 2013). The chicken genome is just over a third the size of a typical mammalian species, being only 1.2 Bbp. Similar to other livestock species the genome sequence has been used to create a variety of public genotyping tools such as the Illumina 60K chip (Groenen *et al*. 2011) and the Affymetrix high-density chip (Kranis *et al*. 2013), and the additional development of company specific arrays.

Opportunities

There are a variety of opportunities afforded to chicken breeding through genomic technologies such as high-throughput genotyping and sequencing. Two such opportunities described in this paper include genomic selection and identification of DNA variations explaining deleterious phenotypes.

Genomic selection

In boiler production, the gains of genome selection are made through the improvement of accuracy of selection, and through the introduction of new traits that could not otherwise be incorporated into the breeding program, rather than reducing generation interval. In Cobb, the

current analytical tool used to estimate genomic breeding values is single-step genomic BLUP (ssGBLUP) (Aguilar *et al*. 2010; Christensen & Lund 2010) using BLUP90IOD (Aguilar *et al.* 2011; Tsuruta *et al*. 2011). This methodology is amenable to our program due to the simple and fast calculation of genomic breeding values.

In order to calculate the impact of genomic selection on our pedigree traits, both traditional and genomic evaluations are computed and compared. The accuracy of each evaluation is determined by correlating the corrected phenotype with the predicted breeding values (either traditional BLUP or ssGBLUP) when the phenotype is not included in the analysis. These estimates of accuracy indicate that the improvement in breeding value accuracy due to genomic selection is highly variable and dependent on the heritability and the number of birds with genotypes for the trait(s) in question.

The key to the improvements in accuracy of breeding values for ssGBLUP is the increase in accuracy of estimated relationships between genotyped individuals. Estimated genomic relationships between full-sibs ranges from 0.266-0.701 with a mean of 0.483 while half-sib relationships range from 0.050-0.547 with a mean of 0.239 (Figure 2).

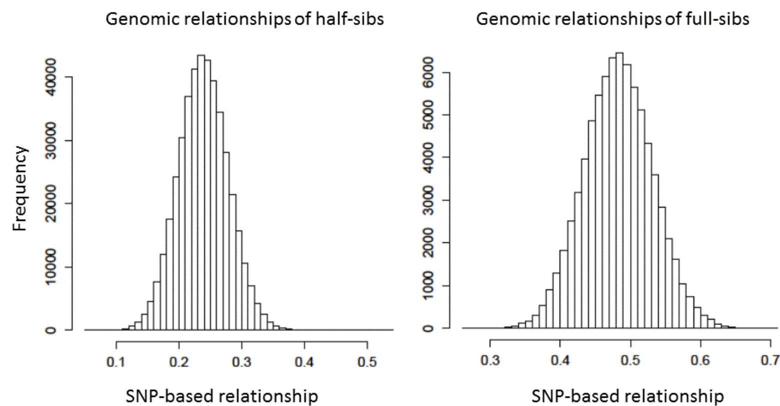


Figure 2. Histogram of genomic relationships among half-**s**ibs (left) calculated from 431778 half sib pair combinations, and full sibs (right) calculated from 78352 full sib pair combinations.

The improvements in accuracy of selection of traits measured in our pedigree program represent the first step in the application of genome selection to a chicken breeding company. We anticipate that the largest gains for genomic selection will be for the incorporation of new traits that can only be measured on chickens outside the pedigree facility (such as disease challenge, and broiler performance in commercial environments as a four-way cross broiler).

DNA variations explaining deleterious phenotypes

High-throughput sequence analyses can be utilised to identify causal or predictive mutations for particular phenotypes. One such effort was toward the identification of causal or predictive alleles for a phenotype specific to one of our pure line breeds. This phenotype was termed 'wiry down', where affected chicks appeared wet and lethargic, and in most cases died soon after hatch. Pedigree analyses of affected families indicated that this phenotype was likely the result of a genetic mutation that occurred in one sire, seven years prior to the phenotype becoming obvious at our hatchery. In order to identify the mutation for this genetic disease, high-throughput genome sequencing was completed on pooled samples representing affected and unaffected individuals. Allele frequencies were compared between pools which highlighted a 10Mb region on chromosome 4 associated with the phenotype. Subsequent fine mapping of this region identified a

single SNP that was 100% predictive of the phenotype. This SNP is now being used to eliminate the condition from our population.

Similar efforts to identify predictive mutations for broiler phenotypes have not all been successful. Some of these efforts have utilised the same pooling approach as above and some have utilised an individual sequencing approach. The incomplete genome sequence, inaccuracies of phenotype recording and the complex nature of some of these phenotypes impact the successful identification of predictive tests for all traits.

Challenges

There are a variety of challenges that impact the utility of genomic tools in Cobb-Vantress.

- The chicken genome sequence is currently incomplete. In spite of the continual improvement of the chicken sequence, it is estimated that the current build is missing ~20% of the total genome (Warren 2014). Some of these missing sequences are due to missing micro-chromosome sequences (9 completely missing, and one other is poorly covered) and approximately 30,000 gaps in the available sequence (W. Warren *pers. comm.*). More importantly for the success of our genomic selection program, this missing sequence is estimated to contain between 5% and 20% of the expressed genes. This presents a difficult challenge in our ability to completely scan the chicken genome for genetic elements contributing the expressions of phenotypes.
- Current sequencing technologies are unable to sequence the GC-rich micro-chromosomes. Therefore tools such as genotyping-by-sequencing, or low coverage genome sequence for use in genomic selection will also fail at scanning the entire genome for contributions toward trait expression.
- The development of 'stable' and multiple-line genotyping tools (like SNP chips) is complicated by the massive allele frequency differences both between pure line populations, and the rapid changes in allele frequencies between generations of the same line.
- Logistics is one of the greatest challenges for the implementation of genomic selection. Examples of obstacles to overcome are:
  - The timing of sampling; genomics is simply not cheap enough to sample and process every chick at hatch; therefore, sampling has to be completed strategically.
  - Genotype processing time; the available time between sampling and selection age for the calculation of genomic breeding values is very short.
  - Sample collection and management; thousands of samples are collected and processed on a weekly basis from a number of pedigree farms both in the US and Europe.

While many challenges exist for implementing genomics, the opportunities and potential gains for a chicken breeding program are large.

REFERENCES

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al*. (2010) *J. Dairy. Sci.*, **93**: 743.

Aguilar I., Misztal I., Legarra A. and Tsuruta S. (2011) *J. Anim. Breed. Genet.* **128**: 422.

Christensen O.F. and Lund M.S. (2010) *Genet. Sel. Evol.* **42**: 2.

Groenen M.A, Megens H.J., Zare Y., Warren W.C., Hillier L.W., *et al.* (2011) *BMC Genomics*. **12**: 274

Hillier L. *et al* (2004) *Nature* 432; 695.

Krani, A., Gheyas A.A., Boschiero C., Turner F., Yu L., *et al*., (2013) *BMC Genomics*, 14; 59.

Tsuruta S., Misztal I. and Strandén I. (2011) *J. Anim. Sci.* **79**: 1166.

Warren W. (2014) XXIV PAG conference proceedings

GENOMIC PREDICTION USING SEQUENCE DATA IN A MULTIBREED CONTEXT

M.S. Lund[1], I. van den Berg[1,2,3] and D. Boichard[2]

[1] Center for Quantitative Genetics and genomics, Department of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark
[2] INRA, UMR1313 GABI, 78350, Jouy-en-Josas, France
[3] AgroParisTech, Paris, France

SUMMARY

Sequence data can potentially increase the reliability of multi breed genomic prediction by containing causative variants or markers in in high linkage disequilibrium (LD) with those. Sequence data does, however, also contain a large number of variants in low LD with the causative mutations, limiting the potential increases in prediction reliability when the full sequence would be used directly for genomic prediction. The objective of this study was to use sequence variants to increase the reliability of multi breed prediction in dairy cattle. First, a simulation study based on real sequence data was carried out to investigate how sequence variants can improve the reliability of across breed prediction. The simulation study used the regression of genomic relationships at causative mutations on genomic relationships at prediction markers to measure the loss in prediction reliability as a consequence of using markers in imperfect LD. It was concluded that it is important to use only variants very close to the causative mutations. In the second part a number of two component Bayesian SNP BLUP models were used, where the first component mainly model variation within the breeds, while the second component model covariance across the breeds. Here, sequence variants selected from a multi breed GWAS for production traits were used as prediction markers in the second component. Different models and selection strategies were compared. Large increases in reliability, up to 0.10, were observed for multi breed prediction using QTL variants compared to within breed prediction using only 50K markers. Our results show that using a selective number of sequence variants can result in large increases in reliability, but careful selection of the variants is essential

INTRODUCTION

The reliability of genomic prediction is highly dependent on the size of the reference population. While for some breeds, for example Holstein, there are large national and international reference populations available, reliabilities in other breeds can be limited due to the smaller size of the reference populations. Smaller breeds can potentially benefit from the large reference populations available for some breeds by multi breed prediction. In practice, however, multi breed prediction only results in substantial increases in reliability compared to within breed prediction when closely related breeds are combined (Lund *et al.,* 2014). One reason for this could be that linkage disequilibrium (LD) is only conserved over short distances across breeds, and therefore, the density of marker chips is insufficient to allow across breed prediction (de Roos *et al.*, 2008). While the markers on the high density (HD) chip are dense enough for across breed prediction, the HD chip did not result in substantial increases compared to the 50K chip. The reason is likely that increasing the density to HD or full sequence does not only add variants closer to the causative mutations, but also variants in low LD with the causative mutations. Unless only variants in complete LD with the causative mutations are used, a loss in prediction reliability occurs (de los Campos *et al.*, 2013). By including QTL variants selected from the sequence, Brøndum *et al.* (2015) found increases in reliability up to 5% for within breed prediction of production traits in dairy cattle. Because LD is conserved over shorter distances across breeds than within breed, such an approach can potentially be more beneficial for across breed and multi breed prediction than for

within breed prediction. Methods for multi breed predictions must be able to 1) capture the genetic variance within breed without introducing noise from private variants or SNP associations only present in one (potentially dominating) breed and 2) capture covariance across breeds by markers in very close LD with causative variants segregating in multiple breeds.

The objective of this study was to use sequence variants to increase the reliability of multi breed prediction in dairy cattle. First, a simulation study based on real sequence data was carried out to investigate how sequence variants can improve the reliability of across breed prediction. Subsequently, sequence variants associated with QTL detected for milk, fat and protein yield were used for multi breed prediction in three dairy cattle breeds.

MATERIALS AND METHODS

For the simulation study, realised sequences on chromosome 1 of 122 Holstein, 27 Jersey, 28 Montbéliarde, 23 Normande and 45 Danish red bulls were used. Causative mutations were randomly sampled from 1,475,541 bi-allelic SNP and indels on chromosome 1, or from all variants with a minor allele frequency (MAF) below 0.10. The number of causative mutations was 10, 50, 100 or 250. Different sets of prediction markers were compared, with all variants from the 50K or HD chip, only the 50K or HD variants closest to each causative mutations, or sequence variants in two 1 Kb intervals on either side of each causative mutation. In the latter scenarios, the distance between intervals and causative mutations varied from 1 base to 1 Mb, and the intervals contained either all variants or only the variants with a MAF of at least 0.10.

For each scenario, two genomic relationships matrices were constructed for each breed and each pairwise combination of breeds, using either the causal loci, or the prediction markers. Genomic relationship matrices were scaled using the allele frequencies computed using the genotypes of all individuals in the genomic relationship matrix. Subsequently, the loss in R^2 was computed following de los Campos et al. (2013):

$$\overline{R}^2_{n+1,y} \leq R^2_{n+1,y}\left[1-(1-b_{n+1,y})^2\right],$$

where, for individual $n+1$, the difference between the prediction ($\overline{R}^2_{n+1,y}$) using markers in imperfect LD with the causative mutations and the prediction ($R^2_{n+1,y}$) if prediction markers were in perfect LD with causative mutations is quantified by the reliability factor (RF) $1-(1-b_{n+1})^2$. The b in the RF is the regression coefficient of the genomic relationships at prediction markers on the genomic relationship markers at the causative mutations. First, b was computed for each individual. Subsequently, RF was computed within replicate, using the b averaged across individuals. Finally, RF was averaged across replicates.

The second part of the study used imputed sequences and deregressed proofs (DRP) for milk, fat and protein yield from 5,852 French Holstein, 5,411 Danish Holstein, 1,203 Danish Jersey and 937 Danish Red bulls. First, bulls genotyped with the 50K chip were imputed to HD. For the French data, this step was performed using Beagle 3.0.0, while for the Danish breeds, IMPUTE2 was used. Subsequent imputation to whole-genome sequence was for all breeds done using IMPUTE2. The reference used for imputation to sequences of the Danish bulls consisted of the bulls in run 4 of the 1000 bull genome project, while for the imputation of the French bulls, a combined French-Danish reference set was used. The latter consisted of 122 Holstein, 27 Jersey, 28 Montbéliarde, 23 Normande and 45 Danish Red bulls.

A number of Bayesian SNP BLUP models were run. As prediction markers, either the 50K markers, or 50K markers in one component and sequence variants selected from a multi breed GWAS (van den Berg *et al.*, these proceedings) were used in a second component. Different selection strategies were compared. Selecting either all variants with a p-value in the multi breed

GWAS below 10^{-10}, 10^{-14} or 10^{-20}, or selecting maximum 1, 10 or 25 variants per intervals of 1, 2 or 10 Mb. Genomic breeding values were estimated using a Bayesian SNP BLUP model. Both single trait models, using a breed effect to account for differences between breeds, and multi trait models, fitting the same trait in different breeds as different correlated traits were used. The single trait models contained a within breed 50K component (ST-WB50K), a multi breed 50K component (ST-MB50K), or multi breed 50K and QTL components (ST-MB50K-MBQTL). The multi trait models contained a multi breed 50K component (MT-MB50K), multi breed 50K and QTL components (MT-MB50K-MBQTL) or a within breed 50K and a multi breed QTL component (MT-WB50K-MBQTL). For all models, marker effects and variance components were estimated using Bayz software, with a MCMC chain of 50,000 iterations, discarding the first 10,000 as burn-in. Reliabilities were estimated as the squared correlation between DRP and GEBV, divided by the mean reliability of DRP in the test population.

RESULTS AND DISCUSSION

RF decreased rapidly when the distance between prediction markers and causative mutations increased. This decrease was larger for across breed prediction than within breed prediction. Figure 1 shows the RF as a function of the distance between causative mutations and prediction markers for across breed prediction. Sequence variants on an interval on a similar distance to the causative mutations as the closest 50K or HD marker resulted in a larger RF when all 50K or HD variants were used, while RF was largest when only the closest 50K or HD markers were used. This shows that, in order to benefit from full sequence data, it is important to use only variants in high LD with the causative mutations, rather than using all sequence variants.

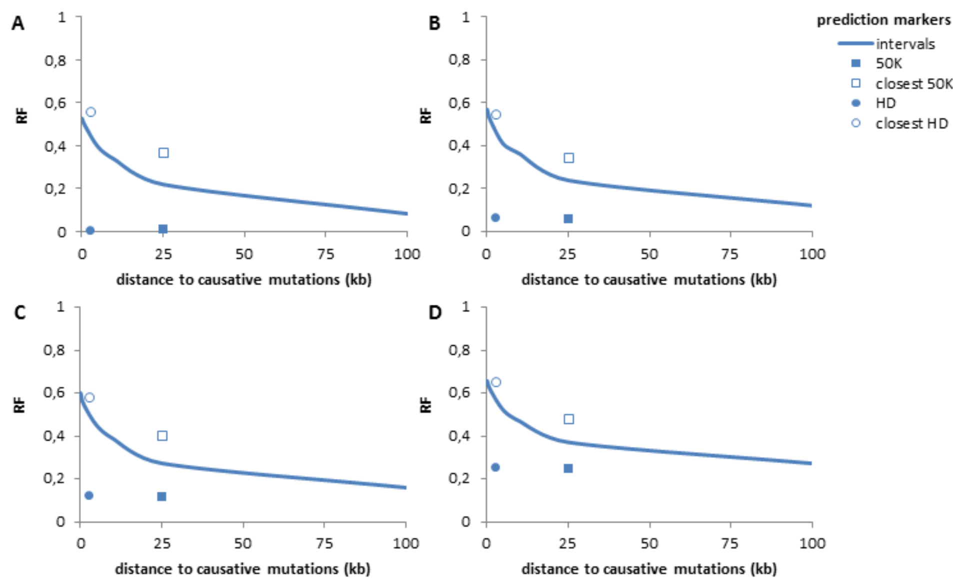


Figure 1. Average across breed reliability factor (RF) computed from intervals or from SNP from the 50K or HD chips, for different numbers (*c*) of causative mutations. A: *c*=10, B: *c*=50, C: *c*=100, D: *c*=250.

Using sequence variants selected from a multi breed GWAS resulted in substantial increases in reliability for all breeds and traits. The reliability was, however, highly sensitive to the set of

prediction markers used. Maximum increases compared to within breed prediction using 50K markers ranged from 0.042 for fat yield in Jersey to 0.105 for milk yield in Jersey. For all breeds and traits, the highest reliabilities were obtained when the number of variants per QTL interval was limited. Selecting many variants per QTL risks the selection of variants that are not in LD across breed, and, thereby, lowers the reliability. While for both Danish and French Holstein, best results were obtained with single trait models, the multi trait models generally resulted in higher reliabilities in Jersey and Danish Red. Because most of the individuals in the data were Holstein, Holstein had a much larger effect on the estimated marker effects than Jersey and Danish Red. Although some QTL are shared across breeds, this is not the case for all QTL, and markers associated with QTL segregating in Holstein but not in the other breeds could introduce noise. A multi trait Bayesian variable selection that would allow different sets of prediction markers to influence different breeds, could potentially lead to larger increases in reliability than those observed here.

Table 1. Scenarios with largest prediction R^2 for each breed and trait. Δ is the difference with within breed prediction using 50K markers. HOLDK = Danish Holstein, HOLFR = French Holstein, JER = Jersey, RDC = Danish Red.

Breed	Trait	50K	Best scenario	Δ
HOLDK	Milk	0.440	ST-MB50K-MBQTL10-25/1	0.087
	Fat	0.475	ST-MB50K-MBQTL20-25/1	0.103
	Protein	0.388	ST-MB50K-MBQTL10-1/1	0.055
HOLFR	Milk	0.327	ST-MB50K-MBQTL14-25/1	0.079
	Fat	0.367	ST-MB50K-MBQTL20-25/1	0.097
	Protein	0.372	ST-MB50K-MBQTL14-25/10	0.065
JER	Milk	0.299	MT-WB50K-MBQTL20-1/10	0.105
	Fat	0.161	ST-MB50K-MBQTL10-10/10	0.042
	Protein	0.219	MT-WB50K-MBQTL20-1/10	0.049
RDC	Milk	0.136	MT-MB50K-MBQTL20-10/1	0.073
	Fat	0.114	MT-MB50K-MBQTL20-25/10	0.075
	Protein	0.093	MT-WB50K-MBQTL14-10/10	0.059

Our results, both from simulation and real data, show that using a selective number of sequence variants can result in large increases in reliability for multi breed prediction in dairy cattle, but careful selection of the variants is essential.

ACKNOWLEDGEMENTS

IB benefited from an Erasmus-Mundus fellowship and a grant by Apisgene, within the framework of the European Graduate School in Animal Breeding and Genetics. This research was supported by center for Genomic Selection in Animals and Plants (GenSAP) funded by The Danish Council for Strategic Research).

REFERENCES

Brøndum R.F., Su G., Janss L., Sahana G., Guldbrandtsen B., Boichard D. and Lund M.S. (2015) *J. Dairy Sci.* **98**:4107.

de los Campos G., Vazquez A.I., Fernando R., Klimentidis Y.C. and Sorensen D. (2013) *PLoS Genet.* **9**:e1003608.

de Roos A.P.W., Hayes B.J., Spelman R.J. and Goddard M.E. (2008) *Genetics* **179**:1503.

Lund M.S., Su G., Janss L., Guldbrandtsen B. and Brøndum R.F. (2014) *Livest. Sci.* **166**:101.

SELECTION PATTERNS FOR HOLSTEIN SIRES IN PRODUCTION-RECORDED HOLSTEIN HERDS WITH DIFFERING FEEDING SYSTEMS

K.L. Macmillan[1], J.M. Morton[2] and J.E. Pryce[3]

[1]Department of Veterinary Science and Agricultural Science, The University of Melbourne, Princes Highway, Werribee, VIC 3030
[2]Jemora Pty Ltd, PO Box 2277, Geelong, VIC 3220
[3]Biosciences Research Division, Department of Economic Development, Jobs, Transport & Resources, Agribio, Bundoora, VIC 3083

SUMMARY

The dataset created for the Feeding the Genes Project (Morton *et al*. 2013) was used to estimate the extent of phenotypic responses from Holstein daughters of Holstein sires in milk yield and composition variables associated with a 50 unit increase in the cow's sires' Australian Profit Ranking (APR). A subset from that dataset comprising 77,144 cows born from 2005 to 2009 was used to compare sire selection patterns in herds with different feeding systems. Australian Profit Ranking values of the cows' sires' varied widely from -303 to +430, including 20% of cows that were daughters of sires with negative APR values. Australian Profit Ranking values of the sires of the enrolled cows were low, with an average of 68 and an average annual rate of increase of 10.1 APR units. The cows in the herds with the greatest reliance on pasture and feeding <1,000kg grain supplement in the bail had the highest average APR of 77 and the highest average annual rate of increase of 13 APR units, whereas the cows in the herds feeding total mixed rations with minimal reliance on pasture had the lowest average APR of 48 and an annual average increase of 7.8 APR units. These results indicate that there are broad ranges in average APR values and in annual average increases in APR, across the different feeding systems. Major differences in sire selection patterns were also observed among the different feeding systems.

BACKGROUND

The APR is a selection index that estimates the relative profitability of different animals and enables ranking of bulls based on the estimated relative profitability of their daughters. Australian Profit Ranking values are calculated using Australian Breeding Values (ABVs); these are based on data derived from herds using a diverse range of feeding systems. Some farmers question the relevance of the APR to herds that have limited or no reliance on pasture grazing, such as those using feedlot systems and feeding total mixed rations. These questions were addressed in the Feeding the Genes Project (Morton *et al*. 2013). The phenotypic changes in yields and milk composition associated with a 50 unit increase in the cow's sires' APR were compared between cows in herds with each of 5 feeding systems.

Although sire selection patterns were known to differ among feeding systems, across herds differing in average yield levels, and in different states and regions, none of these differences have been quantified. Most of the 505 herds enrolled in the Feeding the Genes Project used feeding systems that combined grazing improved pastures with varied amounts of grain supplements fed in the bail during milking.

The objective of the current study was to utilise the dataset created for the Feeding the Genes Project to assess whether Holstein sire selection patterns varied by feeding system, and to measure any differences in genetic trends associated with these systems.

MATERIAL AND METHODS

All herds in which at least 50 Holstein cows calved in 2011 were selected from the Australian Dairy Herd Improvement Scheme (ADHIS) database. Letters were sent to herd managers asking them to complete a simple herd data questionnaire to identify their feeding system. In total, 505 herds provided data suitable for analysis, and cow and lactation data for these herds were obtained from ADHIS. The original dataset included 250,857 lactations for Holstein cows born from 2002 to 2009. The 77,144 Holstein cows, with identified sires, that had at least one enrolled lactation commencing between 2008 and 2011 were selected for the current study. Each cow's sires' APR was as estimated on 20th August, 2012. These cows were in 438 herds and were born from 2005 to 2009. Each herd was classified into one of five feeding systems for each year from 2008 to 2010 (Table 1).

Table 1. Details of five feeding systems used to classify 438 Holstein herds enrolled in the Feeding the Genes Project.

System name	System no.	System description	% of cows
Low bail feeding	1	≤1000kg grain supplement/cow/lactation	11.6
Moderate to high bail feeding	2	>1000kg grain supplement/cow but did not use feed pad or mixer wagon	59.0
Partial mixed ration (PMR)	3	Part of the ration was fed on a feed pad using a mixer wagon with pasture for at least 9 months /year	17.1
Hybrid	4	Pasture for 2-8 months of the year and entirely on a feed pad with a mixer wagon for some periods	6.3
Total mixed ration (TMR)	5	Cows are usually fed a total mixed ration with less than 1month/year on pasture	6.1

RESULTS

The most common feed management system was moderate to high bail feeding. It included 66% of the 1885 herd-years enrolled in the current study and 59% of the enrolled cows (Table 1).

Sires that each had at least 300 daughters had APR values ranging from -303 to +430, and 96% of study cows were daughters of sires with APR values between -200 and +200. The distribution of these cows by sire APR is shown in Figure 1. Overall, 20% of cows were by sires with negative APR values. This varied from 16% for cows in low bail feeding herds (System 1; see Table 1) to 26% of cows in TMR herds (System 5). By comparison, 34% of all cows were from sires with APR values of at least 120, varying from 34% in low bail feeding herds to only 20% in TMR herds. While 22% of cows born from 2005 to 2007 had sires with negative APRs, 16% and 14%, respectively, of cows born in 2008 and 2009 had sires with negative APR values. Mean APR values and selected ABVs within each feed management system are shown in Table 2. Daughters in low bail feeding herds had a sire average APR of 77 compared to only 48 for cows in TMR herds. Similarly, the average annual change in sire APR increased by 13.0 units per year for cows in the low bail feeding herds, but by only 7.8 units per year for cows in the TMR herds (Table 2).

Sire ABVs also reflected differences in sire selection criteria between feeding systems. Whereas the mean of sire milk volume ABV was only 259 litres for cows in the low bail feeding herds, this increased to an average of 311 litres for cows in the TMR herds. This was in contrast to the declining trends (from low bail feeding to TMR) in means of sire ABVs for protein and milk

fat yields (Table 1).

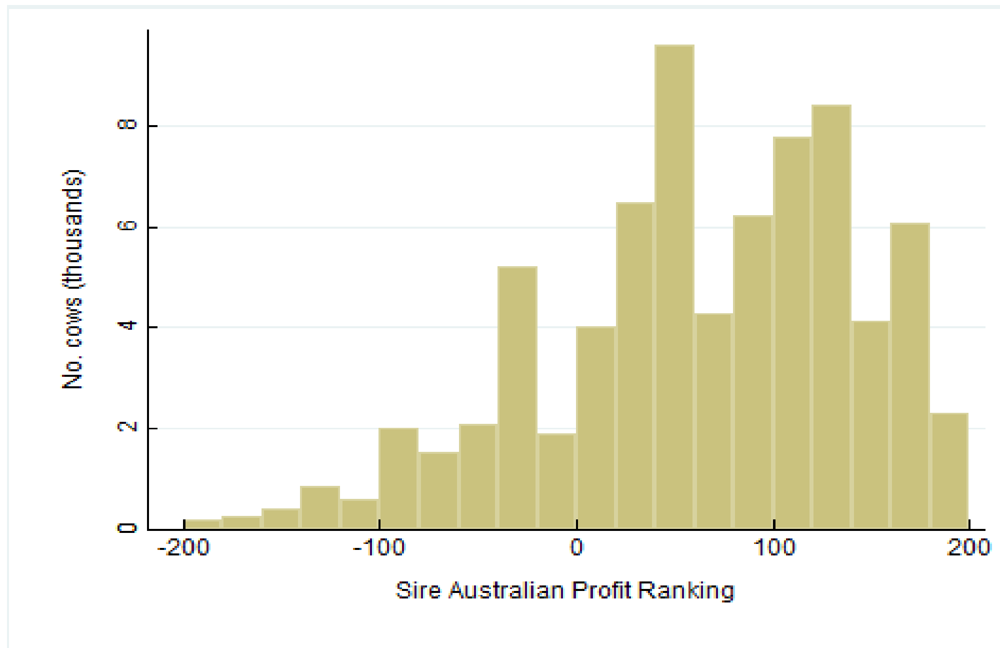


Figure 1. Distribution of Holstein cows by APR of their sire for cows with sires with APRs between -200 and +200.

Table 2. Mean (± SD) sire APRs and annual increases in sire APR and their mean sire ABVs among cows in herds with differing feeding systems.

Genetic variable	Low bail	Mod/high bail	PMR	Hybrid	TMR	Pooled
APR						
Mean	77(85)	71(85)	62(84)	62(89)	48(86)	68(86)
Change/year	13.0	10.0	8.8	12.1	7.8	10.1
ABV						
Milk volume (l)	259(461)	289(472)	302(470)	296(482)	311(459)	289(470)
Protein yield (kg)	9(17)	8(18)	7(18)	7(19)	6(19)	8(18)
Milk fat yield (kg)	9(12)	8(12)	5(12)	7(12)	2(12)	7(12)

DISCUSSION AND CONCLUSION

There was a wide range in cow's sires' APR values and almost 20% of the cows were daughters of sires with negative APR value. This pattern of sire selection indicated that there had been limited emphasis on sire APR by many herd owners; this consequently contributed to an average APR of only 68 and an annual average APR increase of 10.1.

Sires with high APR values were sometimes selected, as 14% of cows were daughters of sires with APR values greater than 160. However, the overall average of 68 could have been expected to be twice as high if herd owners had only selected sires with high APR values. Mean sire APR values differed markedly between feeding systems. The lower APR values for cows in the TMR herds may partly reflect the greater emphasis on milk volume ABV and less emphasis on milk fat and protein yield ABVs in herds with this feeding system, as the APR is negatively influenced by milk volume ABV and positively influenced by protein yield ABV. This may also have contributed to the slower annual rate of increase in the average APR of cows in these herds.

A key finding from the Feeding the Genes Project was that a 50 unit increase in cow's sires' APR is associated with increases in phenotypic protein and milk fat yields as well as profitability in all feeding systems (Morton *et al*. 2013). The average increases in the yields of milk volume, protein and milk fat measured in cows in the TMR herds was approximately double the APR associated increases for cows in the low bail feeding herds (110 versus 56 litres; 5.1 versus 2.6 kg protein; and 5.7 versus 2.6 kg milk fat/cow/lactation for TMR versus low bail feeding, respectively; Morton *et al*. 2013). These yield improvements were calculated to increase milk profit by $46/cow/lactation for a 50 unit increase in APR for cows in the TMR herds compared to $22 for cows in the low bail feeding herds (Morton *et al*. 2013). If greater use had been made of sires that had APR values greater than 160 (as was the case with 14% of cows) to achieve an average increase in APR of approximately 100 units higher that the pooled average of 68 units of APR (Table 2), the overall increase in milk profit would have averaged over $50/cow/lactation across the herds enrolled in the study. It would have been slightly less for cows in low bail feeding herds ($45), but almost double this for cows in TMR herds ($92).

In conclusion, these results showed that greater use of sires with higher APR values could increase milk profit for herds in each of the 5 feeding systems with the greatest increases occurring in herds that relied less on pasture as the major source of nutrients.

REFERENCES

Morton J., Brightling P. and Little S. (2013) Feeding the Genes – Interactions between feeding system and APR and ABVs in dairy cows, pp 1-107. Australian Dairy Herd Improvement Scheme, Melbourne.

MILDLY PENALIZED MAXIMUM LIKELIHOOD ESTIMATION OF GENETIC COVARIANCES MATRICES WITHOUT TUNING

Karin Meyer

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

SUMMARY

A scheme for penalized estimation of genetic covariance matrices free from tuning – using default settings for the strength or penalization – is described and its efficacy is demonstrated by simulation.

INTRODUCTION

Estimates of genetic covariance matrices, $\mathbf{\Sigma}_G$, are known to be afflicted by substantial sampling errors, increasing markedly with the number of traits considered. 'Regularization', i.e. modification of estimators to reduce sampling variation at the expense of a small, additional bias, has been advocated to obtain estimates closer to the population values. An early suggestion by Hayes and Hill (1981, 'bending') has been to shrink the canonical eigenvalues, λ_i, i.e. the eigenvalues of $\mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_G$ (with $\mathbf{\Sigma}_P$ the phenotypic covariance matrix), towards their mean. As shown by Meyer and Kirkpatrick (2010), the analogue in a maximum likelihood framework is to maximize the likelihood subject to a penalty proportional to the variance among the estimates of λ_i. Neither authors provided guidelines on how to determine the amount of shrinkage to be applied. While cross-validation techniques allow estimation of so-called 'tuning factors', this proved laborious and only moderately successful (Meyer 2011).

A simple alternative is to apply a mild, default penalty which, while not providing maximum benefits, will yield stable estimates and worthwhile reductions in 'loss', i.e. the average deviations of estimates from population values. This is similar to the concept of weakly informative priors, which is gaining popularity in Bayesian estimation (e.g. Gelman 2006). This paper demonstrates the reductions in loss achievable using a default penalty on canonical eigenvalues.

PRIORS AND PENALTIES

For a given prior distribution of some function of the parameters to be estimated, we can obtain a corresponding penalty as minus the logarithmic value of the pertaining probability density.

Shrinking canonical eigenvalues towards their mean, $\bar{\lambda}$, by applying a quadratic penalty, $\mathcal{P} \propto \sum_i (\lambda_i - \bar{\lambda})^2$, implies a Normal distribution, $\mathrm{N}(\bar{\lambda},\sigma^2)$, with σ^2 the variance of λ_i. This gives penalty

$$\mathcal{P}_N = \frac{1}{2}\left[q\big(\log(\sigma^2)+\log(2\pi)\big) + \frac{1}{\sigma^2}\sum_{i=1}^{q}\big(\lambda_i-\bar{\lambda})^2\big)\right] \quad (1)$$

with q denoting the number of traits. Similarly, assuming a log-Normal distribution, the penalty is obtained by substituting $\log\lambda_i$ and $(\sum_i \log\lambda_i)/q$ for λ_i and $\bar{\lambda}$ in (Eq. 1). Earlier results showed that for such a prior it was advantageous to penalize both $\log\lambda_i$ and $\log(1-\lambda_i)$ (Meyer 2011). We use $\mathcal{P}_L$ to denote the penalty obtained by summing contributions for both, with the same variance σ^2.

A more flexible alternative is a Beta distribution, $\mathrm{B}(\alpha,\beta)$, with scale parameters α and β allowing for a wide range of shapes. For $\alpha,\beta > 1$, the distribution is unimodal. In Bayesian estimation, $\nu = \alpha+\beta$ is interpreted as effective sample size, i.e. the number of 'observations' added by the prior. A Beta distribution with mode equal to $\bar{\lambda}$ can be specified as $\alpha = 1+(\nu-2)\bar{\lambda}$ and $\beta = 1+(\nu-2)(1-\bar{\lambda})$, for $\nu > 2$. This allows us to quantify the degree of belief in the prior through the single parameter ν. For $m = \nu-2$ and $\Gamma(\cdot)$ denoting the Gamma function, the penalty for the Beta distribution is

$$\mathcal{P}_\beta = q\left[\log\Gamma\big(\nu\big) - \log\Gamma\big(1+m\bar{\lambda}\big) - \log\Gamma\big(1+m(1-\bar{\lambda})\big)\right] + m\left[\bar{\lambda}\sum_{i=1}^{q}\log\lambda_i + (1-\bar{\lambda})\sum_{i=1}^{q}\log(1-\lambda_i)\right] \quad (2)$$

*AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

As for $\mathcal{P}_L$, $\mathcal{P}_\beta$ involves functions of $\log\lambda_i$ and $\log(1-\lambda_i)$. The strength of penalization for all three penalties is regulated by the parameter σ^2 or ν. In contrast to previous formulations employing a tuning factor, this lends itself to attempts of direct estimation by maximizing the penalized likelihood with respect to this parameter (de los Campos 2013; pers. comm.).

SIMULATION STUDY

Records for $q = 9$ traits were sampled from multivariate normal distributions, assuming a balanced paternal half-sib design comprised of $s = 100{,}400$ or 1000 sire families of size 10. Population values for 72 scenarios, selected to represent an extensive range of possible – including unusual or 'difficult' – cases were obtained by combining 12 sets of heritabilities with 6 correlation structures. The variance among population values for canonical eigenvalues ranged from 0 to 0.099 (mean 0.046) on the original and 0 to 2.504 (mean 0.989) on the logarithmic scale. Restricted maximum likelihood estimates of genetic and residual ($\mathbf{\Sigma}_E$) covariance matrices were obtained fitting a simple animal model with means as the only fixed effects, for the three types of penalties described above. Penalties were applied using the same, default 'strength parameter' for all cases, $\sigma^2 = 0.02$ to 0.1 for $\mathcal{P}_N$, $\sigma^2 = 0.5$ to 2.0 for $\mathcal{P}_L$ and $\nu = 2.5$ to 10 for $\mathcal{P}_\beta$. In addition, σ^2 or ν were estimated from the data by evaluating points on the profile likelihood and employing a quadratic approximation to determine its maximum. In doing so, parameter estimates were constrained to the interval [2.001,50] for ν and [0.001,10] and [0.01,25] for σ^2 for $\mathcal{P}_N$ and $\mathcal{P}_L$, respectively. A total of 500 replicates were carried out for each case. For each sample, the loss in estimates was determined as (for $X = G, E$ and P)

$$L_1\left(\mathbf{\Sigma}_X, \hat{\mathbf{\Sigma}}_X\right) = \mathrm{tr}\left(\mathbf{\Sigma}_X^{-1}\hat{\mathbf{\Sigma}}_X\right) - \log\left|\mathbf{\Sigma}_X^{-1}\hat{\mathbf{\Sigma}}_X\right| - q \tag{3}$$

with $\mathbf{\Sigma}_X$ the matrix of population values, $\hat{\mathbf{\Sigma}}_X$ the corresponding estimate, and $\mathbf{\Sigma}_P = \mathbf{\Sigma}_G + \mathbf{\Sigma}_E$. The **P**ercentage **R**eduction **I**n **A**verage **L**oss due to penalization was then evaluated as

$$\mathrm{PRIAL} = 100\left[1 - \bar{L}_1\left(\mathbf{\Sigma}_X, \hat{\mathbf{\Sigma}}_X^\nu\right) / \bar{L}_1\left(\mathbf{\Sigma}_X, \hat{\mathbf{\Sigma}}_X^0\right)\right] \tag{4}$$

with $\hat{\mathbf{\Sigma}}_X^\nu$ and $\hat{\mathbf{\Sigma}}_X^0$ the penalized and unpenalized estimates of $\mathbf{\Sigma}_X$, and $\bar{L}_1(\cdot)$ the average loss over replicates. In addition, the mean reduction in unpenalized likelihood due to penalization (from its maximum for unpenalized estimates), $\Delta\mathcal{L}$, was calculated.

RESULTS

Our main goal of penalized estimation is to reduce the loss in estimates of $\mathbf{\Sigma}_G$. The rationale for shrinking canonical eigenvalues towards their mean or mode is that this reduces sampling variation by 'borrowing strength' from the estimate of $\mathbf{\Sigma}_P$, which typically is estimated much more accurately than either of its components, $\mathbf{\Sigma}_G$ and $\mathbf{\Sigma}_E$ (Hayes and Hill 1981). Hence, loosely speaking, we attempt to redress the balance in partitioning skewed by sampling error. This implies that we expect the estimate of $\mathbf{\Sigma}_P$ to remain more or less unchanged. Too stringent penalization can result in reduced or even negative PRIAL for any of the covariance matrices estimated. In particular, a negative PRIAL for $\mathbf{\Sigma}_P$ represents a strong warning signal for over-penalization.

The distribution of PRIALs (higher values are better) across the 72 scenarios for $\mathbf{\Sigma}_G$ and $\mathbf{\Sigma}_P$ for two sample sizes is summarized in Figure 1. Central dots display mean values. Values on the x-axis are the fixed values for σ^2 and ν used, except for 'E' which denotes use of the estimated value. For clarity of scale, 5 values for $\mathbf{\Sigma}_G$ less than −60, occurring for $\mathcal{P}_L$, 'E' and $s = 100$, are omitted. PRIALs for $\mathcal{P}_N$ were modest, but positive throughout for $\mathbf{\Sigma}_G$. Except for the smallest, fixed value for σ^2, i.e. the most stringent penalty, there was little evidence for adverse effects on estimates for any of the cases. Estimating σ^2 yielded marked improvements, especially for the larger sample size, for cases with low variance among the population values for λ_i, i.e. the cases which matched the prior. As there were few of the latter, however, mean PRIALs achieved remained quite low.

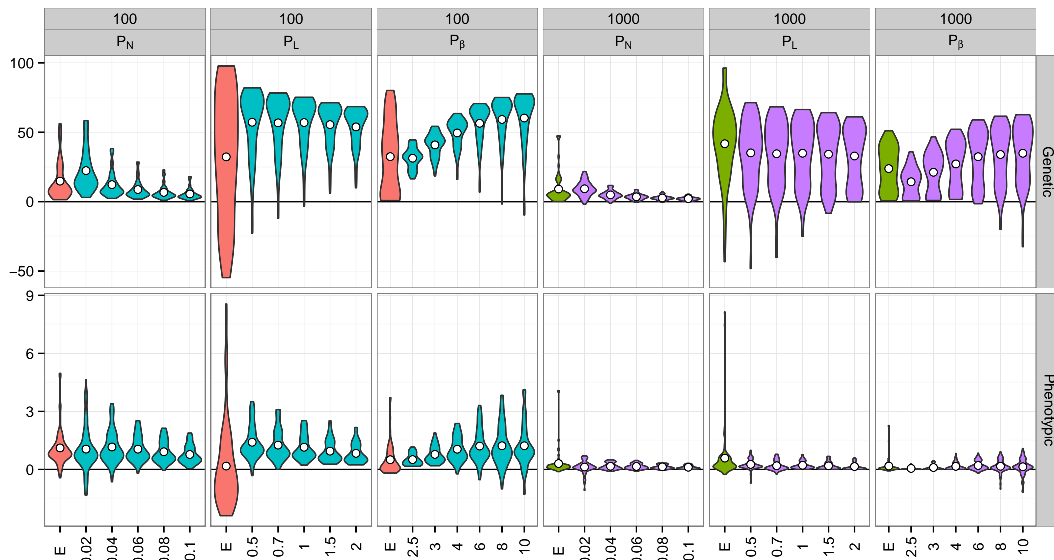


Figure 1. Distribution of Percentage reduction in average loss for genetic and phenotypic covariance matrices (for s=100 and 1000 sire families).

As reported previously (Meyer 2011), penalizing $\log \lambda_i$ yielded substantially larger PRIALs. Using a fixed value of $\sigma^2 = 2$ proved to be a safe default for a mild penalty with mean PRIALs for $\mathbf{\Sigma}_G$ as high as 54, 44 and 33% for $s = 100, 400$ and 1000, respectively. Lower values for σ^2 resulted in increasing numbers of unfavourable cases. Attempts to estimate σ^2 for $\mathcal{P}_L$ failed in a substantial proportion of replicates for a number of cases, with estimates close to the lower boundary. As this was set at 0.01, it resulted in far too stringent penalization. This held in particular for the smallest sample size, suggesting that this was, in part at least, attributable to insufficient information. Additional analyses (not shown) estimating separate values of σ^2 for two the parts of $\mathcal{P}_L$, involving $\log \lambda_i$ and $\log(1 - \lambda_i)$, respectively, reduced the incidence of problem cases, but, on the whole, was not satisfactory either.

A similar pattern emerged for a penalty based on the Beta distribution. However, for $\mathcal{P}_\beta$ an estimate for ν close to its lower boundary at 2.001 was equivalent to virtually no penalization. Hence, there were no negative PRIALs due to over-penalization. Yet, overall there was little advantage in estimating ν compared to a default value for a mild penalty. Means and minimum values for PRIALs and the corresponding decrease in likelihood (from it's unpenalized maximum) for selected values of ν are given in Table 1. Results for fixed ν identified little adverse effects for any cases or sample sizes for values up to about 6. Average changes in likelihood were small, especially when considering that for $q = 9$ traits there were 90 covariance components to be estimated. As shown in Table 1, repeating analyses with minimum values for ν of 2.5 and 4 increased PRIALs by a few percent compared to corresponding results for a fixed ν, but at the price of marked additional effort.

DISCUSSION

Penalized estimation provides a powerful mechanism to improve estimates of genetic covariance matrices by reducing sampling variation. Large improvements can be obtained if population parameters approximately match the assumed prior distribution on which the penalty is based. In practice, however, true values are unknown and it is important that the procedure chosen is robust, i.e. unlikely to result in worse estimates. While estimation of the strength of penalization is possible in principle, it is computational demanding and may not be particularly advantageous or even successful.

Table 1. Mean and minimum PRIAL for estimates of genetic (Σ_G), residual (Σ_E) and phenotypic (Σ_P) covariance matrices and change in log likelihood ($\Delta\mathcal{L}$) for penalty $\mathcal{P}_\beta$

ν^a	100 sires				400 sires				1000 sires			
	Σ_G	Σ_E	Σ_P	$\Delta\mathcal{L}$	Σ_G	Σ_E	Σ_P	$\Delta\mathcal{L}$	Σ_G	Σ_E	Σ_P	$\Delta\mathcal{L}$
						Mean						
F2.5	31.3	36.4	0.5	-0.27	21.9	15.6	0.1	-0.08	14.3	7.6	0.1	-0.03
F4.0	49.5	48.2	1.0	-1.25	37.5	23.3	0.4	-0.47	27.2	12.6	0.2	-0.22
F6.0	56.5	51.5	1.2	-2.52	43.6	26.3	0.5	-1.06	32.4	14.7	0.2	-0.55
F8.0	59.2	52.6	1.2	-3.72	45.7	25.5	0.5	-1.68	34.0	12.7	0.2	-0.92
E2.0	32.4	16.7	0.5	-1.25	31.9	11.8	0.4	-0.56	23.8	7.5	0.2	-0.27
E2.5	46.2	46.5	0.9	-1.35	38.8	24.2	0.5	-0.60	28.2	13.1	0.2	-0.29
E4.0	54.2	52.7	1.2	-2.01	46.7	29.3	0.6	-0.84	33.1	16.3	0.3	-0.41
						Minimum						
F2.5	16.2	8.7	0.2	-0.47	0.9	1.3	0.0	-0.23	0.2	0.4	0.0	-0.11
F4.0	16.0	22.5	-0.2	-1.98	4.0	5.0	0.0	-1.13	1.4	1.7	0.0	-0.69
F6.0	6.7	23.4	-0.5	-3.91	-0.4	9.4	-0.6	-2.41	-2.0	3.4	-0.2	-1.69
F8.0	-1.7	8.4	-1.0	-5.69	-17.1	13.2	-0.9	-3.72	-20.2	-1.3	-1.0	-2.67
E2.0	0.3	-3.1	-0.2	-5.21	0.5	-0.7	-0.1	-1.96	0.5	-0.2	-0.1	-1.18
E2.5	18.1	26.5	0.0	-4.99	14.5	7.1	-0.0	-2.01	3.4	3.0	-0.0	-1.17
E4.0	16.2	32.5	0.0	-5.38	15.8	9.8	0.1	-2.02	5.1	3.9	-0.1	-1.19

[a]Effective size, F: fixed value, E: estimated with this minimum value

Results show that a penalty encouraging shrinkage of canonical eigenvalue lends itself to a scheme using a default strength parameter to impose a mild penalty. Assuming a Beta distribution provides a more flexible prior than a Normal or log-Normal distribution. Moreover, the resulting penalty has an intuitive parameter – the so-called effective sample size – regulating its stringency. A value of $\nu = 4$ to 6 yielded worthwhile reductions in loss without (non-negligible) negative PRIALs or substantial changes in likelihood and can be recommended for routine use. Additional computational requirements are small, but derivatives of the penalty may be needed. These are readily obtained by parameterising analyses to canonical eigenvalues and elements of the corresponding eigenvectors (Meyer and Kirkpatrick 2010), but this may have less favourable convergence rates than the standard parameterisation. Alternative penalties, e.g. to shrink genetic correlations towards their phenotypic counterparts may be preferable in this respect, and equally suited to default penalties (Meyer 2014).

CONCLUSIONS

Maximum likelihood estimation subject to a penalty can markedly reduce sampling variation of estimates, and should be applied routinely in multivariate analyses involving more than a few traits.

ACKNOWLEDGEMENTS

Work was supported by Meat and Livestock Australia grant B.BFG.0050

REFERENCES

Gelman A. (2006) *Bayesian analysis* **1**:515.
Hayes J.F. and Hill W.G. (1981) *Biometrics* **37**:483.
Meyer K. (2011) *Genet. Sel. Evol.* **43**:39.
Meyer K. (2014) In *Proc. Tenth World Congr. Genet. Appl. Livest. Prod.* Paper No. 217.
Meyer K. and Kirkpatrick M. (2010) *Genetics* **185**:1097.

EFFICIENCY OF A TACTICAL PHENOTYPING STRATEGY FOR MULTI-STAGE SELECTION

C Massault and J.H.J van der Werf

School of Environmental and Rural Science, University of New England, Armidale, NSW 2351

SUMMARY

Accurate genetic evaluation relies on measurements, which can be difficult to achieve for some economical important traits (hard and /or costly to measure). We developed a strategy that can select an optimised subset of animals to phenotype based on pedigree relationship, prior information (previously phenotyped animals) and diversity to maximise genetic gain under inbreeding and cost constraints.

We simulated a two-stage two-trait selection scenario for a small population of 10 paternal half-sib families of size 10 (pilot study). One trait was phenotyped for all animals (parents and selection candidates) and the second trait was phenotyped on only a selected set of 20 selection candidates based on a prior decision on phenotyping (stage 1). Phenotyping decisions were made either based on maximizing diversity of the set chosen to be phenotyped (DIVERSITY) or simply based on breeding values at stage 1 (MERIT). After phenotyping, the second stage selection of animals as parents for the next generation was based on optimum contributions. The DIVERSITY strategy was most useful when there was limited prior information about the Mendelian sampling term of predicted breeding value. When parents of selection candidates have not been phenotyped, DIVERSITY does not provide any advantage over truncation selection (MERIT). However, when sires or both parents have been previously phenotyped, DIVERSITY resulted in higher genetic gain for similar level of inbreeding. From this study, we conclude that an optimized phenotyping strategy can have potential long term benefits in breeding programs but more work is needed to investigate under which conditions benefits are largest.

INTRODUCTION

Trait measurement provides the necessary information to perform accurate genetic evaluation, whether it is based on phenotype on the animal itself or on its relatives. However, it can be costly and/or difficult to achieve trait measurement for a large number of animals (e.g. carcass traits, methane emission). Reducing the number of animals phenotyped is a simple and efficient way to cut cost and/or allow economically important traits to be part of the selection criterion, but the question is how phenotyping costs can be reduced with minimal impact on genetic gain.

Various efforts to manage cost of phenotyping have been made over the years. The first attempt to manage measurement was made by Robertson (1957) who proposed a theory to optimise the family size in a progeny testing breeding program by optimising the product of expected selection differential and accuracy. Wade and James (1990) developed a theory to manage the cost of testing while limiting a reduction in genetic gain. They mainly optimised the proportion of selection candidates to be phenotyped. More recently, Okeno et al. (2014) found that using knowledge on previously estimated breeding values was better than phenotyping randomly selected animals and that a phenotyping 80% of the animals provides the same gain as when all animals were phenotyped. Previous studies were therefore mainly concerned with determining an optimal proportion of animals to be phenotyped, but did not give any insight about which particular individuals should be measured.

In a previous study, we developed a phenotyping strategy for a single trait measured on an optimised subset of animals with no prior information on the candidates to be considered (Massault *et al.* 2013). A set of individuals was phenotyped that maximised the information and

thus the accuracy of genetic evaluation. While this strategy proved to be efficient, it does not reflect a practical breeding program where multi-trait and multi stage selection is common and prior information on candidates exists in the form of estimated breeding values.

In the current study, we present a phenotyping strategy that uses prior knowledge for the case of two-stage two-trait selection. We use a small paternal half-sib population structure and extreme parameters to assess the potential usefulness of such a strategy.

MATERIALS AND METHODS

Population simulation. To explore the efficiency of our selection criterion, we simulated a small pedigree of 10 paternal half-sib families, comprising 10 offspring each. We simulated genetic and environmental values for 2 traits with both a heritability of 0.3 and a phenotypic variation σ^2_P of 100 (e.g. body weight and feed efficiency) with a correlation between traits of r_A = 0.5 and r_E = 0.25. Traits 1 and 2 have economic value of 0.01 and 0.1 respectively (the most important trait being the one with restricted phenotyping). We therefore have a breeding objective G of 0.01 * EBV_{Trait1} + 0.1 * EBV_{Trait2} (EBV = Estimated Breeding Value). We used selection in stage 1 to determine which selection candidates to phenotype for Trait 2, and offspring were selected to become parents at stage 2. All animals were measured for Trait 1 before stage 1 selection. We had three different scenarios for prior information on Trait 2; NOT2, where no parents of selection candidates have been phenotyped for Trait 2; ST2, where sires have been phenotyped for Trait 2 and PT2, where both parents have been phenotyped for Trait 2. These scenarios differ in the amount of information known about the between and within family components of EBV, and hence the correlation in EBV among relatives We calculated EBVs at stage 1 using multi-trait BLUP based on all available information on both traits. EBVs in stage 2 were calculated using BLUP after phenotyping 20% of selection candidates for Trait 2. We used optimum contribution selection at stage 2 (Sonesson and Meuwissen, 2000), where contribution of animals to the next generation are optimised and balanced with diversity:

$$Optimum\ Contribution = \boldsymbol{x'G} + \lambda_2 \boldsymbol{x'Ax}$$

where $\boldsymbol{G}$ is a vector of breeding values, $\boldsymbol{x}$ a vector of contribution to the next generation and $\boldsymbol{A}$ the numerator relationship matrix. We used 6 different values for λ_2 (0, -10,-100, -1000, -9999). We then compared the genetic gain (**x'G**) for the same level of inbreeding (F = **x'Ax** / 2) between different phenotyping strategies.

Selection criteria for phenotyping (stage 1 selection). We used 2 strategies to select 20% of selection candidates to phenotype: DIVERSITY, where animals are selected for phenotyping based on merit as well as diversity, and MERIT, where animals are selected simply based on merit (highest EBVs). We also simulated an ALL strategy where all selection candidates were phenotyped. We propose a selection criterion based on the average (EBVs) of 'would-be' phenotyped animals and their genetic diversity:

$$Selection\ criterion = \boldsymbol{x'G} + \lambda_1 \boldsymbol{x'Ax}$$

where $\boldsymbol{G}$ is the vector of expected breeding value of phenotyped animal, $\boldsymbol{x}$ a vector indicating for each animal 1/**n** (number of phenotyped animals) when phenotyped or 0 otherwise and $\boldsymbol{A}$ the numerator relationship matrix. The first term reflects the average breeding values of phenotyped animals while the second term reflects the diversity between phenotyped animals. We use the extreme value of -9999 for λ_1 for DIVERSITY. Note that the MERIT scenario where the phenotyped set is chosen based on merit alone is equal to setting λ_1 to zero, while a scenario where diversity was the overriding criterion equal one where λ_1 = 9999. The optimality of the

result will also depend on the importance of genetic diversity at the final selection stage, hence we also varied λ_2. We simulated 100 replicates.

Differential algorithm. To determine which subset of animals is best to phenotype, we used a differential algorithm (DE, Storn and Price 1997). The DE creates 16 subsets of animals to phenotype (solutions). Each subset of phenotyped animals is then evaluated using a selection criterion (described above) and the solutions are ranked. The DE creates a challenger for each of the 16 solutions by crossing-over and mutating solutions (i.e. a different set of animals to be phenotyped). If the challenger performs better, the current solution is discarded and the challenger enters the pool for the next generation of solutions. Subsequently challengers are evaluated with the selection criterion over another 50,000 generations. At the end, the DE will have evolved to a best (or at least close to best) set animals to phenotype, for a given λ_1 value.

RESULTS AND DISCUSSION

Genetic gain and diversity. Fig. 1 shows the possible selection points at stage 2 (plotting genetic gain (**x'G**) and level of inbreeding F (**x'Ax**/2)) for the two selection options at stage 1, with MERIT (i.e. λ_1 = 0) and DIVERSITY(λ_1 = 9999) for NOT2, ST2 and PT2 scenarios. ALL always out-performs the 2 other strategies, as expected, due to the fact that all selection candidates have been phenotyped. The ALL strategy gave by far the highest gain under the same level of inbreeding at stage 2 selection. The DIVERSITY strategy performed generally better than the MERIT strategy. The advantage is not so important in the case of NOT2, but DIVERSITY is significantly higher when phenotyping sire (ST2) and both parents (PT2). The performance of DIVERSITY increases as the level of information on parents increases.

Between and within family information. The difference seen between DIVERSITY and MERIT can be explained by the additional information given by phenotyping parents of selection candidates. In the case of NOT2, where no parents are phenotyped, there is no information other than the selection candidates own phenotype for Trait 1. EBVs from sibs are lowly correlated (correlation = 0.06, see Stanish and Taylor (1983) for calculation) which reduces the co-selection of relatives. Therefore, in this case, a phenotyping strategy that emphasises diversity is not really advantageous. MERIT distributed phenotypes to best animals across 8 different families (Fig 2.A) while DIVERSITY phenotyped 2 good animals for each family (Fig 2.B) resulting in DIVERSITY slightly better than MERIT. On the other hand, when sires have been phenotyped for Trait 2, sibs will have a higher correlation (0.40) among their EBVs. Selection on merit would emphasize family selection, which becomes restrictive at stage 2 selection (unless λ_2 = 0) and the distributing phenotypes over more families will allow more emphasis on within family selection at stage 2. Fig. 2.C shows that MERIT allocated phenotypes for the 5 best families while Fig 2.D shows that DIVERSITY phenotyped 2 good animals per family. This permits a relatively high genetic gain and also maximises the diversity and show the benefit of choosing specific individuals to phenotype rather than a random proportion to maintain diversity. The same principle applies when both parents have been phenotyped and the additional information brought more information on the selection candidates itself as each dam has a single progeny.

Further work. The results showed in this study proved that in selecting animals for phenotyping there is an advantage to emphasize diversity of the set to be measured. We concluded that phenotyping good animals across a larger number of families resulted in higher genetic gain than phenotyping the best animals of few families for same level of inbreeding. An optimal solution is likely to be less extreme than the DIVERSITY strategy, hence it is important to find the λ_1 value that optimises the subset of animals to phenotype that allows an optimal solution between merit and diversity at the second selection stage. It is also pertinent to explore the benefit of an optimal strategy, over a long period of time (e.g. 10 years of selection) and vary the parameter values such as trait heritabilities, genetic correlation and economic weights.

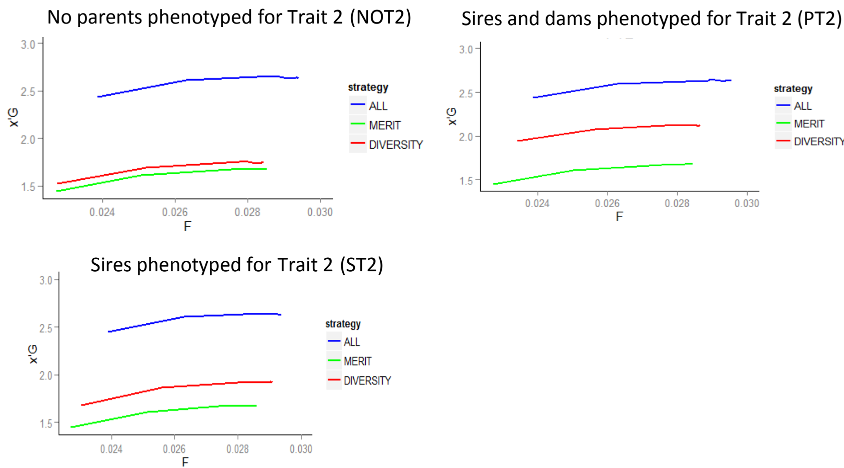


Fig 1. Genetic gain (x'G) versus diversity (F, level of inbreeding) plot for 3 different measurement scenarios NOT2, ST2 and PT2 with the three phenotyping strategies ALL, MERIT and DIVERSITY

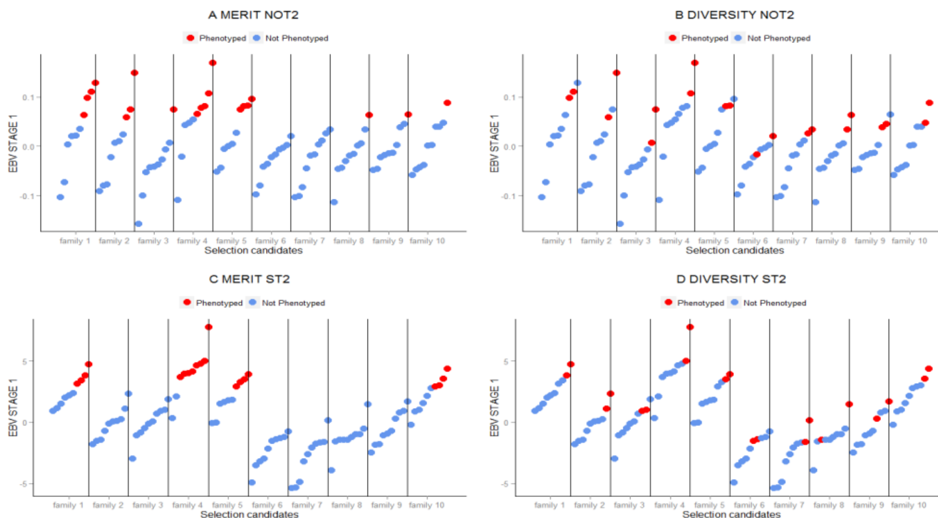


Fig 2. Phenotyped individuals in one replicate with no information on Trait 2 for MERIT(A) and DIVERSITY (B) and when sires are phenotyped for Trait 2 for MERIT (C) and DIVERSITY (D) . Individuals are classified by EBV at stage 1 and families.

REFERENCES

Massault, C, Kinghorn, B., van der Werf, J.H.J (2013) *Proc. Assoc.Advmt. Anim. Genet.* **20**: 416
Okeno, T. O., Henryon , M., Sørensen, A. C. (2014) *Proc. WCGALP*. **10**
Robertson, F.W. (1957) *Biometrics* **13**: 442
Sonesson, A.K., Meuwissen, T.H.E. (2000) *Genetics Selection Evolution* **32**: 231
Stanish, W.M., Taylor, N. (1983) *The American Statistician*, **37**: 221
Storn, R., Price, K. (1997) *Journal of global optimization*, **11**: 341
Wade, C.M., James, J.W. (1990) *Proc. Assoc. Advmt. Anim. Genet*. **8**:115

A COMMERCIAL COMPARISON OF EWE BREEDS FOR REPRODUCTION, WOOL AND LAMB GROWTH

K. P. Ransom[1], F. D. Brien[2] and W.S. Pitchford[2]

[1] 19 Neale St. Bendigo, Vic. 3550.
[2] School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy SA 5371

SUMMARY

Sheep farmers are actively seeking unbiased information on the performance of sheep breeds and bloodlines within breeds to help improve their overall profitability. To help answer this question the Elmore Field Days Inc ran a comparison to determine the merit of five ewe genotypes for prime lamb and wool production from 2009 to 2014. Each of the five genotypes were represented by 42 ewes randomly selected from three properties. The ewes were joined annually to terminal sires for prime lamb production and run together as one mob except at lambing; there were six opportunities to lamb, the first as ewe lambs. Ewe genotypes compared were the Border Leicester x Merino cross (BL x Mo), local Merinos from northern Victoria and three dual purpose Merinos, Centre Plus Merinos, the Dohne Merino and the South African Meat Merino (SAMM). The local Merinos produced the heaviest fleeces and Centre Plus Merinos the finest. The BL x Mo and SAMMs had the highest reproduction but lighter and coarser fleeces.

INTRODUCTION

What is the best ewe genotype to use for a combination of prime lamb and wool production? To help answer this question the Elmore Field Days Inc ran a trial from 2009 to 2014 to compare the merit of five alternative sheep genotypes in the northern Victorian environment at Elmore. This paper reports covers the results of five adult years of ewe and lamb body weights, condition scores, reproduction data and key wool measurements. Additional data for other traits were available for some years and are also reported.

MATERIALS AND METHODS

Ewes were run on the Elmore Field Days site 3 km east of Elmore in northern Victoria from January 2009 to October 2014. The rain at the locality is winter dominant with a long term average of 466mm per year. Sheep grazed on annual pastures growing between late autumn and spring and dry pasture residues and crop stubbles over the summer. Summer storms in some years provided extra green feed from dryland lucerne and green summer weeds.

Five ewe genotypes were each represented by 42 ewes. Each genotype group was randomly selected from three properties, with 14 ewe lambs per property after an allowance for culling. The ewe lambs were fed a high-quality diet to reach a joining weight in late February 2009, when they were joined to White Suffolk rams with a further five annual joinings to either White Suffolk or Poll Dorset rams.

The ewe genotypes were (i) BL x Mo. - Border Leicester x Merino cross ewes, the most common prime lamb mother in northern Victoria. (ii) Merino LV - Loddon Valley Merinos, the second most common prime lamb mother in northern Victoria; based on Peppin bloodlines with some influence from South Australian bloodlines and typical of many medium Merinos. These local ewes were compared to three dual purpose Merinos that have been recently introduced to the district. (iii) CP Merino - Centre Plus Merino, from Central West NSW. (iv) Dohne - The Dohne Merino is a dual-purpose breed developed in South Africa. In 2008 the Dohne breed was in the early stages of introduction to Australia from South Africa. Two properties supplied F_2 ewe lambs while the third property supplied F_3 ewes. (v) SAMM - The South African Meat Merino is a

dual-purpose sheep originally bred in South Africa. In 2008 the SAMM breed was in the early stages of introduction to Australia. Two properties supplied F_3 ewe lambs while the third supplied a mix of F_2s and F_3s. This report covers the five adult years of body weights, condition scores, reproduction data and key wool measurements. Additional data were available for some years.

Lambing time varied from year to year, from April (autumn) to August (late winter) as ram introduction varied from 1 November and 26 February. Ewes were pregnancy scanned about 90 days after the rams were introduced and assigned as 'dry' or carrying a single or twin. They were divided into their breed groups immediately prior to lambing and run together again from lamb marking. Ewes were inspected twice daily during lambing and assistance was only given when needed. Individual lambs were not identified with their dam at lambing. Instead ewe udders were inspected at lamb marking and weaning and each ewe was classed as 'wet' or 'dry' or 'lambed and lost' when linked to scan information.

Shearing was in early October year and wool mid-side samples for wool quality characters were taken about 3 weeks before shearing. Ewes were scored using standard industry guidelines by two experienced operators for greasy wool colour in two years and for wrinkle (neck and body) once at three years of age. Fleece rot was scored before shearing after a normal season in September 2010 and in April and September 2011, after the wettest summer on record with 611 mm of rain over 5 months. Lamb growth was assessed by live weights before sales and the proportion that would go to slaughter in the first draft. Lambs were weaned at 12 to 14 weeks and sold when a commercial draft reached a minimum live weight of 46 kg; except in the poor spring of 2013 when the weight was reduced to 42 kg. The average age of the first batch at marketing was 21 weeks. Carcase measurements were available for 4 slaughter batches of lambs totalling 460 lambs over three years. Underweight lambs were carried over the summer in two years.

Wool and lamb returns per ewe were calculated each year using average prices over the previous 12 months. Wool prices differences reflected fibre diameter. Lamb returns per ewe were calculated from lambing percentage, lamb live weight, dressing percentage and skin value.

Statistical analyses. A linear mixed model was fitted to reproduction, wool, body weight and condition score traits. Fixed effects fitted within the model included year (confounded with ewe age), ewe genotype and the interaction between year and ewe genotype (which was almost never significant). For ewe body weight and condition score whether or not they lambed at 12 months was included and also the interaction between 12-month-lambing and year. Random effects included property of origin, the interaction between property and year, and ewe to account for repeated measures across years. For wool traits the year by property interactions were not significant and removed from the model. For traits recorded only once (e.g. broken mouths at the end of the trial) the model only included the fixed effect of ewe genotype. For lamb carcass traits, the model included fixed effects of year of birth, sex (ewe, wether) and ewe genotype. Interactions between fixed effects were not significant and removed from the model. Least significant differences (LSD) are shown where appropriate.

RESULTS AND DISCUSSION

Wool and ewe body weight. The local Merino produced the heaviest fleeces and the SAMM the lowest. The Centre Plus Merino had the finest wool, with local Merinos similar to Dohne, SAMM being coarser and the BL x Mo the coarsest (Table 1). SAMM and BL x Mo ewes maintained greater condition than the others, with the local Merino the lowest.

Reproduction and lamb performance. There were no breed differences in fertility (ewes pregnant per ewe joined) but there were substantial differences in fecundity (litter size) and so lambs weaned (Table 2). BL x Mo and SAMM were highest, Centre Plus intermediate and local Merino and Dohne lowest. The 36% difference in lambs marked between the SAMMs and Australian Merinos in this study contrasted with a South African study where the difference was

12% between SAMMs and wool focused Merinos (Cloete 2003). The differences between Dohnes and wool focused merinos were similar in both studies. Heterosis may have had a small effect on the F_2 and F_3 SAMM and Dohne results as maternal heterosis between Australian Merino strains averaged 8% lambs marked in a study of F_1s by Mortimer *et al* (1997).

These analyses clearly show the trial had sufficient numbers to validly test sheep reproduction, weights and wool production with the limited resources of farmer research organisations and a team of dedicated volunteers with professional help. They are a reliable guide to the reproductive performance of the flocks of origin as the property of origin variance was very small and several studies including Allden (1979) and Gunn *et al* (1995) have indicated the nutrition of the young ewe, from a foetus to weaning, has zero or small long term effects on subsequent reproduction under commercial farm conditions.

Lambs from SAMM ewes were the heaviest, then BL x Mo, Centre Plus and Dohne intermediate and the local Merino lowest (Table 3). There were no differences in the GR fat measure after adjustments for carcase weight.

Table 1. Ewe live weight and condition score at joining for the five adult lambings from 2010 to 2014 and wool productions for the five adult shearings from 2010 to 2014.

Ewe Breed	Ewe weight, fleece free at joining (kg)	Condition score at joining (score 1-5)	Greasy fleece weight (kg)	Clean fleece weight (kg)	Fibre diameter mean (µm)	Fibre diam. Coefficient of variation (%)	Greasy wool colour (score 1-5)
BL x Mo	79.6^{b}	4.34^{d}	5.5^{bc}	3.9^{bc}	30.3^{d}	20.3^{c}	3.4^{c}
Merino LV	63.6^{a}	3.55^{a}	6.6^{d}	4.8^{d}	20.9^{b}	18.6^{b}	2.4^{a}
CP Merino	70.7^{a}	3.77^{b}	6.0^{c}	4.1^{c}	19.0^{a}	16.1^{a}	2.4^{a}
Dohne	70.6^{a}	4.05^{c}	5.1^{b}	3.5^{b}	20.8^{b}	17.8^{ab}	2.9^{b}
SAMM	79.7^{b}	4.38^{d}	4.1^{a}	2.6^{a}	24.4^{c}	17.0^{ab}	3.3^{c}
LSD	8.1	0.17	0.5	0.4	1.3	1.1	0.2

^{abc} *Ewe breed means within columns with different superscripts differ significantly (P<0.05).*

Table 2. Ewe reproduction characters for the five adult lambings from 2010 to 2014.

Ewe breed	Scanned in lamb	Fetuses scanned per pregnant ewe	Fetuses scanned per ewe joined	Lambed & lost per ewe lambing	Lambs born per ewe joined	Lambs marked per ewe joined
BL x Mo	0.94	1.67^{b}	1.58^{b}	0.04^{a}	1.50^{c}	1.32^{c}
Merino LV	0.91	1.42^{ab}	1.29^{a}	0.11^{b}	1.16^{a}	0.96^{a}
CP Merino	0.93	1.59^{b}	1.48^{b}	0.07^{a}	1.38^{b}	1.16^{b}
Dohne	0.90	1.40^{a}	1.26^{a}	0.15^{b}	1.20^{a}	0.96^{a}
SAMM	0.97	1.65^{b}	1.61^{b}	0.05^{a}	1.56^{c}	1.32^{c}
LSD	0.08	0.18	0.18	0.07	0.09	0.06

^{abc} *Ewe breed means within columns with different superscripts differ significantly (P<0.05).*

Industry application. Reproduction, lamb growth, wool and easy care characters are all highly relevant to improving profitability, but no single genotype exceled in all compartments. The estimated returns from wool and lamb are shown in Table 4. The local Merino and Centre Plus had the greatest wool returns whereas BL x Mo and SAMM had the greatest lamb returns.

When wool and meat were combined the SAMM and Centre Plus were equivalent and the BL x Mo was close. However, returns per hectare from the BL x Mo and SAMM would be reduced when accounting for their higher feed intake, due to higher number of lambs reared and heavier ewes. However indications are that dual purpose Merinos with good wool, reproduction and lamb growth are likely to be the most profitable alternative in this Elmore environment. Systems analyses using bio-economic models such as GrassGro are needed to fully investigate whole farm profitability.

Table 3. Lamb live weights, growth rates and dressing percentage.

Ewe Breed	Weight at Marking (kg)	Weight in spring, before any sales (kg)	Weight gain, Winter-Spring (g/day)	Percent in 1st slaughter batch	Weight gain over Summer (g/day)	Dressing percent
BL x Mo	19.4^{c}	47.7^{c}	261^{d}	67.4^{c}	156^{d}	47.2^{bc}
Merino LV	18.2^{a}	44.3^{a}	241^{a}	45.3^{a}	142^{b}	46.4^{a}
CP Merino	18.1^{a}	46.1^{b}	258^{c}	55.8^{b}	150^{cd}	46.7^{a}
Dohne	18.7^{b}	46.3^{b}	255^{b}	58.4^{b}	133^{a}	47.1^{b}
SAMM	18.8^{b}	48.5^{d}	274^{e}	72.0^{d}	146^{bc}	47.4^{c}
LSD	0.4	0.5	3	4.0	7	0.3

[abc] *Ewe breed means within columns with different superscripts differ significantly (P<0.05).*

Table 4. Scores for skin wrinkle and fleece rot and the financial returns from wool and lambs.

Ewe Breed	Wrinkle, neck and body (score 1-5)	Fleece rot 8Sept2010 (score 1-5)	Fleece rot 18Apr2011 (score 1-5)	Fleece rot 6Sept2011 (score 1-5)	Wool returns per ewe ($/ewe)	Lamb returns per ewe ($/ewe)	Total wool and lamb returns per ewe ($/ewe)
BL x Mo	1.2^{a}	1.5	1.6^{a}	2.0^{a}	$19.94	$154.54	$174.48
Merino LV	2.5^{c}	1.3	2.4^{bc}	3.2^{b}	$52.96	$102.32	$155.28
CP Merino	2.3^{c}	1.2	2.3^{b}	2.7^{b}	$49.67	$130.04	$179.71
Dohne	1.6^{b}	1.4	2.9^{c}	3.0^{b}	$38.70	$108.68	$147.39
SAMM	1.1^{a}	1.6	2.2^{ab}	2.6^{ab}	$23.17	$157.08	$180.25
LSD	0.2	0.6	0.6	0.6			

[abc] Ewe breed means within columns with different superscripts differ significantly (P<0.05).

ACKNOWLEDGEMENTS

This project was funded and conducted by the Elmore Field Days Inc. together with The Campaspe Prime Lamb Producers Group with professional help from three organisations and six product support sponsors. Australian Wool Innovation provided funds for statistical analyses.

REFERENCES

Allden, W. G. (1979) *Aust J Agric Res*. **30**: 939-948.

Cloete S.W.P., Cloete J., Durand A. and Hoffman L. (2003) *Sth. Afr. J. Anim. Sci.* **33** 223.

Gunn, R.G., Sim, D.A. and Hunter, E.A. (1995) *Anim. Sci.* **60**, 223.

Mortimer, S.I. and Atkins, K.D. (1997) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **l2**:404

REPRODUCTIVE PERFORMANCE OF HOLSTEIN AND JERSEY HEIFERS AND COWS IN A PASTURE-BASED SYSTEM IN SOUTH AFRICA

C.J.C. Muller[1], J.P. Potgieter[2], S.W.P Cloete[1,2] and J.A. Botha[1]

[1]Western Cape Department of Agriculture, Research & Technology Development Services, Directorate: Animal Sciences, Private bag X1, Elsenburg 7607, South Africa;
[2]Department of Animal Sciences, University of Stellenbosch, Stellenbosch 7600, South Africa

SUMMARY

The fertility in dairy cows is a major issue, as several studies suggest declines in the reproductive performance of dairy cows over the past 20 years. Little information is available on the comparative performance of South African Holstein (H) and Jersey (J) cows under similar feeding and management conditions. In this paper, the reproductive performance of H (n=120) and J (n=126) heifers and lactating cows on mostly kikuyu pasture are presented. Cows were supplemented with the same concentrate mixture after milking at 7 kg per cow per day. On average (±s.d.), Jersey heifers were inseminated earlier (P<0.05) than H heifers at 15.4±2.1 and 16.1±2.3 months of age, respectively. A higher (P<0.05) ratio of J heifers were inseminated for the first time by 15 months of age than H heifers (0.49 vs. 0.29). Fertility traits, calving to first service (CFS), first service within 80 days post partum (FS<80d), and cows confirmed pregnant within 100 days post partum (PD100d) for H and J cows were 88±26 and 78±29 days (P<0.01), 0.44 and 0.62 (P<0.01) and 0.31 and 0.51 (P<0.05) respectively. The interval from calving to conception differed (P<0.05) between breeds, being 119±61 and 138±62 for J and H cows respectively Results are consistent with other studies showing a higher conception rate in J cows compared to H cows. The poorer reproductive performance of H could be probably attributed to a greater potential for milk production. Further studies are foreseen comparing the production performance and efficiency of H and J cows under this feeding regime.

INTRODUCTION

The declining fertility of dairy cows has recently become a major issue in most of the main dairy producing countries in the world. In most countries selection is aimed at improving milk production performance and conformation traits. Studies have shown that the declining reproductive performance of dairy cows may be associated to an increasing proportion of North American H sires in national dairy herds (Auldist, *et al*. 2007; Buckley, *et al*. 2003). Because of this decline in fertility in Holstein cows, producers are considering using other breeds, or in some cases, doing crossbreeding to improve traits such a fertility. In South Africa, the J breed is becoming very popular, especially in pasture-based areas. The breed is also increasingly being used in crossbreeding programmes in countries with seasonal pasture-based production systems mostly because of its perceived better reproductive performance in comparison to H cows (Auldist *et al*. 2007). However, Washburn *et al*. (2002a) found in a survey in the USA, an unexpected close similarity between H and J herds for services per conception or conception rate. This differed from earlier work by Fonseca *et al*. (1983) who reported significant differences in first service conception rate, i.e. 72% for J and 49% for H. Washburn *et al*. (2002b) later found that J cows, when managed in the same herd over three years, had a higher conception rate than H cows, being 59.6% *vs*. 49.5%. In contrast, Prendiville *et al*. (2011) found no significant differences in reproductive efficiency between Holstein-Friesian and J cows in a seasonal pasture-based management system in Ireland. Breed differences in the fertility of South African heifers were demonstrated in a survey involving 10721 H heifers in 11 herds and 2349 J heifers in 5 herds

(Muller *et al*. 2014a). Because of a lack of a national data base for insemination or service records and pregnancy check results for dairy cows in South Africa, calving interval (CI) is at present being used as an indicator for fertility. Genetic parameters have been estimated for calving interval (CI) for dairy breeds (Mostert *et al*. 2010). Phenotypically, a small difference in CI was shown between H and J cows, being 398±68 and 389±64 days, respectively. Heritability estimates for CI were low, being 0.022±0.006 and 0.026±0.004 for J and H, respectively, albeit in agreement with other analyses. Mostert *et al*. (2010) found that genetic trends for CI showed an upward curve since 1980, amounting to 1.25 and 0.50 days per year for H and J cows, respectively. Muller *et al*. (2014b) showed that herd (presumably an indicator of managerial and inseminator skills) had the largest effect on the standard of reproduction management in H cows. The aim of the paper is to compare the reproductive performance of H and J heifers and cows under the same feeding and management conditions in a pasture-based feeding system in South Africa.

MATERIAL AND METHODS

Location and Animals. This paper was based on an on-going breed comparison study being conducted at the Elsenburg Research Farm of the Western Cape Department of Agriculture. Elsenburg is situated approximately 50 km east of Cape Town in the winter rainfall region of South Africa. The area has a typical Mediterranean climate with short, cold, wet winters and long, dry summers. Holstein and J cows have been managed since 2003 as one herd. Cows in milk were supplemented with a commercial concentrate mixture being fed after each milking twice a day for a total of 7 kg per day regardless of milk yield and lactation stage. Cultivated pasture consists mainly of kikuyu grass (*Pennisetum clandestimum*) being irrigated during summer. Cows were on kikuyu pasture during most of the year. Pasture was further supplemented during winter with a pasture replacement mixture consisting of lucerne hay, oat hay and soybean meal providing at least 15% CP on an “as is” basis. Fresh drinking water was freely available at all times.

Data recording. Cows were routinely checked and treated by a veterinarian for retained placentas and uterine infections within the first 10 days after each calving. From 40 days after calving, cows were checked for signs of heat and if active, a tail-marker was put on each cow to facilitate heat detection. Cows not showing signs of reproduction activity at this stage were treated according to a standard hormonal programme. Heat detection was done on a daily basis. Cows were inseminated from about 60 days after calving. Heifers born from these cows were put in a heifer-service group once they reached 13 months of age and were checked for reproductive activity. Heifers were serviced when showing clear signs of being in heat. The reproductive performance of heifers and cows was determined based on service dates and the results of pregnancy detection by rectal palpation by a veterinarian at least 45 days after the last service. Reproductive traits determined for cows were the interval (number of days) from calving to first service (CFS), number of services per conception (SPC), interval from calving to conception (DO), whether first service occurred within 80 days post partum (FS<80d), whether cows became pregnant from first service (PDFS) or within 100 (PD100d) or 200 days (PD200d) after calving. Reproduction traits determined for heifers were age at first service (AFS), whether first insemination of heifers was before 15 months of age, conception age of heifers and whether heifers became pregnant before 15 months of age as well as age at first calving (AFC). Categorical traits were scored as 1 for no and 2 for yes.

Statistical analyses. Reproductive traits for heifers and cows were compared between breeds by analysis of variance using SAS. Records within breeds were used as random replicates. For categorical traits, frequency tables and Chi-square tests were used to determine whether response is independent of breed. Significance was declared at $P<0.05$.

RESULTS AND DISCUSSION

Results from the analysis of variance comparing the reproductive performance of H and J heifers and cows are reported in Table 1. Jersey heifers were inseminated earlier (P<0.05) than H heifers, i.e. at 15.4±2.1 and 16.1±2.3 months of age resulting in a higher ratio (P<0.05) of J heifers inseminated for the first time by 15 months of age. The interval CFS was shorter (P<0.05) for J cows in comparison to H cows, being 78±29 vs. 88±27 days, respectively. This resulted in a higher ratio (P<0.05) of J cows being inseminated within 80 days after calving than H cows, i.e. 0.61 vs. 0.44 respectively. While the number of services per conception for J cows only tended (P=0.09) to be less than for H cows, the interval from calving to conception was shorter (P<0.01) for J cows in comparison to H cows, being 119±61 *vs*. 139±62 days respectively. Although average values for some traits were acceptable, large variations were observed as indicated by high standard deviations. The coefficients of variation ranged from 31 to 51% for CFS and DO respectively. The distribution of the number of DO records is shown in Figure 1. The DO interval of more than 100 days is exceeded in 70 and 50% of lactations for H and J cows respectively.

Table 1. Analysis of variance mean (±s.d) estimates of the reproductive performance of Holstein and Jersey heifers and cows in a pasture-based feeding system (AI = artificial insemination; AFC = age at first calving; FS = first service; CFS = calving to first service; DO = days open; DIM = days in milk)

	Heifers			Cows	
Variables	Holstein	Jersey	Variables	Holstein	Jersey
Number of records	120	126	Number of lactations	326	325
Age first service (m)	16.1^{a}±2.3	15.4^{b}±2.1	Lactation number	2.31±1.44	2.56±1.51
First service <15m	0.29^{a}	0.49^{b}	Interval CFS (days)	88^{a}±27	78^{b}±29
AI's per conception	1.86±1.30	1.77±1.08	FS<80 DIM	0.44^{a}	0.61^{b}
Pregnant first service	0.54	0.56	Services/conception	2.19±1.41	1.98±1.32
Conception age (m)	17.5^{*} ±2.9	16.8^{*} ±2.8	Pregnant FS	0.41	0.48
AFC (m)	26.5±2.9	26.1±2.9	Interval DO (days)	139^{a}±62	119^{b}±61
AFC <24m	0.20	0.28	Pregnant <100 DIM	0.31^{a}	0.51^{b}
AFC <27m	0.64	0.69	Pregnant <200 DIM	0.85	0.87

[a,b]Values with different superscripts differ at P<0.05; *Values differed at P=0.07

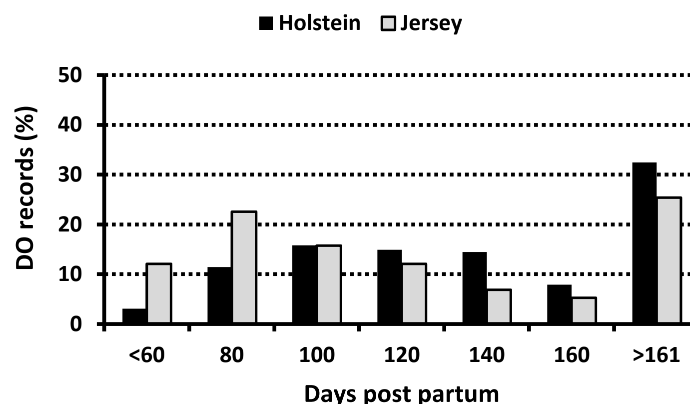


Figure 1. The distribution of the number of records for interval from calving to conception (DO) for all Holstein and Jersey cows

Results are consistent with other studies showing a higher conception rate in J cows in comparison to H cows. The poorer reproductive performance of H could possibly be attributed to a greater potential for milk production. According to an Australian survey (Little, 2003), the observed level of reproductive performance would suggest management problems for both breeds in this study. The 100-day-in-calf rate for H cows was 31% while for J cows 51% was achieved. Mackey *et al*. (2007) reported that in 19 Holstein-Friesian dairy herds in Ireland, fertility performance was generally poor, with the interval to first service being 84.4±35.4 days and the first insemination success rate 40.6±0.68%. The 100-day in-calf rate was 46.0±0.68% and the CI 404±65 days. Growth rate and fertility of heifers are important traits affecting age at first calving and lifetime performance (Cooke *et al*. 2013). More emphasis should be put on the lifetime performance of dairy cows, i.e. total production per day of life from birth, as this would have a greater economic and environmental benefit (Wathes *et al*. 2014).

CONCLUSION

This study reported breed differences in reproduction performance between H and J heifers and cows. Results are consistent with other studies showing a higher conception rate in J cows compared to H cows. Although a larger proportion of J heifers were inseminated before 15 months of age, age at first calving was the same for H and J heifers probably indicating a lack in inseminator proficiency. A larger ratio of H heifers calved down past 27 months of age. First insemination after calving was earlier for J cows compared to H cows, while a higher first service success rate resulting in more J cows confirmed pregnant by 100 days post partum. This translated to fewer days open which should reduce calving interval by approximately 16%. Although J heifers and cows showed a better fertility, a general improvement in reproduction management is required in both breeds. Farmers recognize the importance of fertility in heifers and cows although not using appropriate indicators. Fertility indicators used in the study and results could be used as benchmarks for South African dairy farmers.

ACKNOWLEDGEMENTS

This research was partly funded by the Western Cape Agricultural Research Trust. The efforts of the management team at the Elsenburg dairy are greatly appreciated.

REFERENCES

Auldist M.J., Pyman M.F.S., Grainger C. and Macmillan K.L. (2007). *J. Dairy Sci*. **90**: 4856.

Buckley F.K., O'Sullivan J.F., Mee J.F., Evans R.D. and Dillon P. (2003). *J. Dairy Sci*. **86**: 2308.

Cooke J.S., Cheng Z., Bourne N.E. and Wathes D.C. (2013). Open Access J. Anim. Sci. **3**: 1. http://dx.doi.org/10.4236/ojas.2013.31001.

Fonseca F.A., Britt J.H., McDaniel B.T., Wilk J.C. and Rakes A.H. (1983). *J. Dairy Sci*. **66**: 1128.

Little S. (2003). The InCalf Book for dairy farmers. Dairy Australia.

Mackey D.R., Gordon A.W., McCoy M.A., Verner M. and Mayne C.S. (2007). *Animal*. **1**: 29.

Mostert B.E., Van der Westhuizen R.R. and Theron H.E. (2010). *SA J. Anim. Sci*. **40**: 156.

Muller C.J.C., Potgieter J.P. and Cloete S.W.P. (2014a). *Proc. 10th WCGALP*. p. 640.

Muller C.J.C., Potgieter J.P., Cloete S.W.P. and Dzama K. (2014b). *SA J. Anim. Sci*. **44**: 54.

Prendeville R., Shalloo L., Pierce K.M. and Buckley F. (2011). *Irish J. Agric. Food Sci*. **50**: 123.

Washburn S.P., Silvia W.J., Brown C.H., McDaniel B.T. and McAllister A.J. (2002a). *J. Dairy Sci*. **85**: 244.

Washburn S.P., White W.J., Green J.T. and Benson G.A. (2002b). *J. Dairy Sci*. **85**: 105.

Wathes D.C., Pollott, G.E., Johnson K.F., Richardson, H. and Cooke J.S. (2014). *Animal*. **8:s1**: 91.

GENETIC DIVERSITY AND POPULATION STRUCTURE OF FOUR SOUTH AFRICAN SHEEP BREEDS

L. Sandenbergh[1,2], S.W.P. Cloete[1,3], R. Roodt-Wilding[2], M.A. Snyman[4], and A.E. Van der Merwe[2]

[1]Directorate Animal Sciences: Elsenburg, Private Bag X1, Elsenburg, 7607, South Africa
[2]Department of Genetics, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa
[3]Department of Animal Sciences, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa
[4]Grootfontein Agricultural Development Institute, Private Bag X 529, Middelburg EC 5900, South Africa

SUMMARY

Prior knowledge of the genetic diversity, extent of linkage disequilibrium (LD) and population structure is necessary to determine the sample size and number of SNPs necessary to ensure sufficient power of detection in genome-wide association studies (GWAS) and genomic prediction. The OvineSNP50 chip was used to genotype Dorper, Namaqua Afrikaner (NA), South African Mutton Merino (SAMM) and 2 flocks of South African Merino to determine the genetic diversity, differences in LD across breeds and population differentiation. The NA samples exhibited the least number of polymorphic loci and was also the least genetically diverse breed tested. The South African Merino samples exhibited high levels of diversity comparable to results of international Merinos. The NA samples exhibited the longest stretches of LD in comparison to the 3 other breeds, while the Merino had the most rapid decay in LD. Dorper and SAMM samples exhibited intermediate LD length in comparison to the 2 aforementioned breeds. A principal component analysis (PCA) indicated 4 distinct clusters in the data representing the 4 breeds. The inclusion of additional SAMM and other Merino-based breed samples may aid in increasing the resolution and clearly defining breeds and subtypes.

INTRODUCTION

Genomic prediction and GWAS rely on sufficient marker coverage of the genome and a representative sample cohort (Goddard and Hayes 2009). Estimates relating to the genetic diversity, extent of LD and population differentiation is vital in selecting representative samples and determining the number of markers required for genomic prediction and GWAS (Goddard and Hayes 2009; Zhang *et al*. 2012; Kijas *et al.* 2014).

The South African Merino is the primary fine wool producing breed in South Africa and is also utilised for meat production. The SAMM was originally developed from the German Merino and has become the major dual-purpose breed in South Africa (Cloete and Olivier 2010; Schoeman *et al*. 2010). The Dorper, a 50-50 composite of the Dorset Horn and Persian breeds, is the major meat producing breed in the country (Cloete and Olivier 2010). The NA is a hardy, fat-tailed sheep indigenous to South Africa and is primarily maintained for conservation purposes (Schoeman *et al.* 2010; Qwabe *et al.* 2013). The breed is considered endangered with <1000 breeding ewes and <20 breeding rams remaining (FAO 2000; Qwabe *et al.* 2013). Although the genetic diversity and population structure of South African sheep breeds have been explored previously using microsatellite markers (Soma *et al*. 2012; Qwabe *et al.* 2013), a fine-scale investigation is necessary to confirm the genetic diversity and the breed structure, and determine the extent of LD for future genomic studies (Kijas *et al*. 2012). The current study used the OvineSNP50 chip to genotype 160 Dorper, NA, South African Merino and SAMM samples to investigate differences in genetic diversity, LD and population differentiation across the breeds and sampling groups.

MATERIALS AND METHODS

Samples and genotyping. The Dorper (n=20), NA (n=20) and SAMM (n=20) samples were obtained from a resource flock on the west coast of the Western Cape Province of South Africa at the Nortier Research Farm. The South African Merino samples were obtained from the resource flocks maintained at Cradock (n=50) and Grootfontein (n=50) in the Eastern Cape Province. Blood samples were obtained through venipuncture of the jugular vein and stored between -20°C and-80°C. Samples were thawed and applied to bloodcards for transport. Genotyping was done with the OvineSNP50 beadchip at GeneSeek Inc. (Lincoln, NE, USA).

Data analysis. GenomeStudio Software v. 1.0 (Genotyping Module, Illumina) was used to call genotypes from SNP intensity data and to ensure the stringency of quality control parameters. The following quality control measures were implemented: >0.25 GenCall score; >0.5 GenTrain score; >0.01 minor allele frequency (MAF); >0.95 call rate and a sample call rate >0.95 across all samples. Samples with more than 10% missing data were excluded. Genotype data that met the quality control criteria were used to determine the number of polymorphic loci and the MAF distribution for the 5 respective sampling groups and an additional group comprising 20 Cradock and 20 Grootfontein Merino samples. The observed heterozygosity and inbreeding coefficient (F_{IS}) was calculated for each group in PLINK v.1.07 (Purcell *et al.* 2007). Allelic richness (Ar) and private allelic richness (Par) was determined using ADZE v. 1.0 (Szpiech *et al.* 2008). As SNP ascertainment bias may inflate LD values, LD was calculated for subsets of SNP data pruned within each breed and across breeds. The --indep-pairwise 50 5 0.5 command in PLINK was used to calculate pairwise LD within a 50 SNP window and remove one SNP from a pair where the LD exceeds 0.5 before moving on 5 SNPs and repeating the procedure. Linkage disequilibrium (r^2) was calculated for all SNP pairs remaining after LD pruning using the --r2 command. A principal component analysis (PCA) was conducted in the R package (R Core Team 2015), adegenet v. 1.4-2 (Jombart and Ahmed 2011) to identify population structure within and between the sampling groups and to identify potential outliers. Equal sample numbers (n=20) from each group were included in the PCA. Loci were pruned across all samples and the MAF cut-off was increased to 0.1 to mitigate the possible effect of SNP ascertainment bias. File formatting was conducted in R, PLINK or PGDspider v. 2.0.8.0 (Lischer and Excoffier 2012).

RESULTS AND DISCUSSION

From the total of 160 samples, 16 samples (2 from the Cradock Merino, 13 from the Grootfontein Merino and 1 from the SAMM sampling groups) were excluded. The remaining samples had an average call rate of 99.72% and 91% (of the total of 54 241) of the SNPs met quality control measures (Table 1). The Merino samples (Cradock and Grootfontein) were polymorphic for approximately 89% of SNPs, while NA samples were polymorphic for only 69% of SNP loci. The Dorper and SAMM samples were intermediate to these values, at 83% and 81%, respectively. The MAF distribution of the Merino, Dorper and SAMM were relatively similar and most loci exhibited MAFs of more than 30%. In contrast, the NA samples exhibited a large number of non-polymorphic loci and an equal distribution in the number of polymorphic loci across the MAF range. The NA samples also had the lowest allelic richness, private allelic richness and observed heterozygosity in comparison to the other 3 breeds (Table 1). These low levels of genetic diversity in the NA have also been observed with the microsatellite-based studies (Qwabe *et al.* 2013) and OvineSNP50 genotype information (Kijas *et al.* 2012).

Table 1. Genetic diversity estimates of the 5 sampling groups and a combination sample consisting of an equal number of Cradock (n=20) and Grootfontein (n=20) Merino samples. NA: Namaqua Afrikaner; SAMM: South African Mutton Merino, n: number of samples, MAF: Minor allele frequency; Pn: Percentage of polymorphic loci; SE: Standard error; Ar: Allelic richness; Par: Private allelic richness; He: Observed heterozygosity; F_{IS}: Inbreeding coefficient.

Sample group	n	Loci with MAF<0.01	Pn	Ar (SE)	Par (SE)	He	F_{IS}
NA	20	11921	69.20	1.75 (0.001)	0.007 (0.0003)	0.28	0.25
Dorper	20	4026	83.55	1.89 (0.001)	0.012 (0.0004)	0.34	0.11
SAMM	19	5174	81.16	1.88 (0.001)	0.012 (0.0003)	0.33	0.12
Cradock Merino	48	1120	87.12	1.99 (0.001)	0.014 (0.0004)	0.36	0.05
Grootfontein Merino	37	1120	84.43	1.99 (0.001)	0.011 (0.0003)	0.35	0.08
Merino (combined)	40	1120	89.01	1.94 (0.001)	0.012 (0.0003)	0.35	0.06

The extent of LD varied according to the manner in which LD pruning was applied to the dataset (Figure 1). The unpruned dataset exhibited LD over longer stretches, while pruning within each breed markedly reduced the LD values between SNPs. A less extreme reduction in the extent of LD was observed when SNPs were pruned across breeds. In all datasets, the Merino, followed by the Dorper and SAMM displayed the most rapid decay in LD. The NA samples had the longest stretches of LD overall. High levels of genetic diversity and LD decay over short distances has been reported for international Merino samples and may be a consequence of the large effective population size and variation maintained within the breed (Kijas *et al.* 2012; 2014)

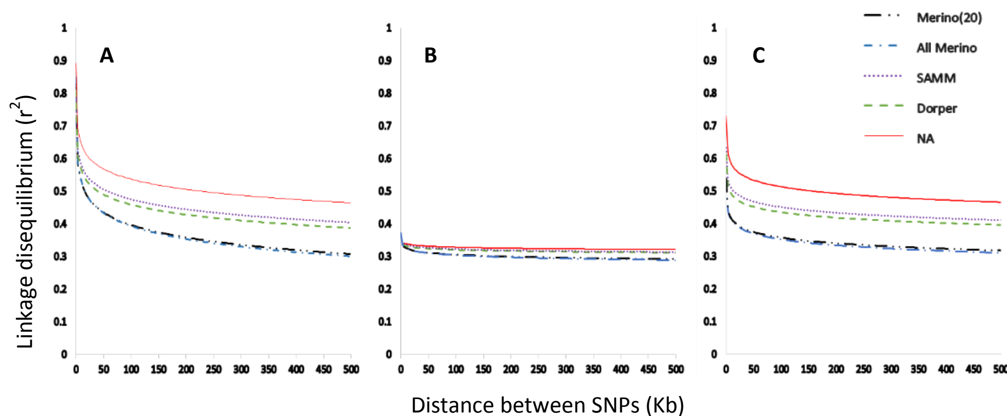


Figure 1. Linkage disequilibrium (r^2) determined for 4 South African sheep breeds prior to pruning SNPs in strong linkage disequilibrium (LD) (A); LD pruning within each breed (B); and LD pruning across all samples (C). Merino (20): Cradock (n=20) and Grootfontein Merino (n=20) samples; All Merino: Cradock (n=48) and Grootfontein Merino (n=37) samples; SAMM: South African Mutton Merino; NA: Namaqua Afrikaner.

The first principal component accounted for 12.29% of the variation in the sample, while the second and third principal components accounted for 7.93% and 6.84%, respectively. Across the first principal component, the NA and Dorper samples clustered separately while substantial overlap was seen between the other sampling groups. The third principal component separated the 4 breeds tested into separate clusters. The Grootfontein and Cradock Merino samples remained clustered together across all principal components. Inclusion of additional Merino samples (48

Cradock Merino, 37 Grootfontein Merino) and the full set of (unpruned) SNPs, resulted in the SAMM samples clustering separately from the Grootfontein and Cradock Merino for all principal components (data not shown).

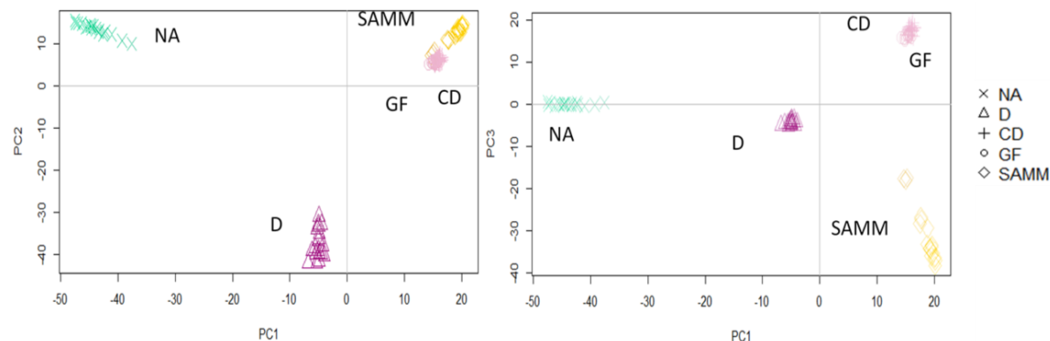


Figure 2. Principal component analysis of 4 South African sheep breeds from 5 sampling groups. (NA: Namaqua Afrikaner; D: Dorper; CD: Cradock Merino; GF: Grootfontein Merino; SAMM: South African Mutton Merino).

The NA samples exhibited large stretches of LD and the least genetic diversity of the breeds tested. Fewer SNPs would therefore be necessary to achieve the same level of coverage of the NA genome than more diverse breeds. Fewer individuals may also be needed to establish a representative sampling cohort for this breed. Despite SNP pruning, the effect of SNP ascertainment bias should still be considered when interpreting whole-genome SNP data from NA as indigenous breeds had limited representation during SNP discovery (Clark *et al*. 2005). The South African Merino samples exhibited high levels of genetic variability and a rapid decay in LD that were comparable to results of international Merino breeds (Kijas *et al*. 2012; 2014). A relatively large sample cohort and a large number of SNPs will be required of future genomic studies to adequately capture all variation contained in this breed. The 4 breeds tested appear to be genetically distinct, however, the inclusion of additional SAMM samples may elucidate the relationship between the SAMM and South African Merino further. Scope exists for further studies that include additional South African sheep breeds, such as the Dormer and Dohne Merino, to clarify the relationship between the South African sheep breeds.

REFERENCES

Clark A.G., Hubisz M.J., Bustamante C.D., Williamson S.H. and Nielsen R. (2005) *Genome Res*. **15**: 1496.
Cloete S.W.P. and Olivier J.J. (2010) The international sheep and wool handbook. Cottle D.J. (ed.), pp. 95.
FAO (2000) World watch list for domestic animal diversity, 3rd ed. Scherf B. (ed.), Rome.
Goddard M.E. and Hayes B.J. (2009) *Nature Reviews Genetics* **10**: 381.
Jombart T. and Ahmed I. (2011) *Bioinformatics* **27**: 3070.
Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., *et al.* (2012) *PLoS Biology* **10**: e1001258.
Kijas J.W., Porto-Neto L., Dominik S., Reverter A., Bunch R., *et al.* (2014) *Anim. Genet*. **45**: 754.
Lischer H.E.L. and Excoffier L. (2012) *Bioinformatics* **28**: 298.
Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., *et al.* (2007) *Am. J. Hum. Genet*. **81**: 559.
Qwabe S.O., Van Marle-Köster E. and Visser C. (2013) *Trop. Anim. Health Prod.* **45**: 511.
R Core Team (2015). R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Schoeman S.J., Cloete S.W.P. and Olivier J.J. (2010) *Livest. Sci*. **130**: 70.
Soma P., Kotze A., Grobler J.P. and Van Wyk J.B. (2012) *Small Rum. Res*. **103**: 112.
Szpiech Z.A., Jokabsson M. and Rodenberg N.A. (2008) *Bioinformatics* **24**: 2498.
Zhang H., Wang A., Wang S. and Li H. (2012) *J. Anim. Sci. Biotech* **3**: 26.

SHEEP PHYLOGEOGRAPHY AND DOMESTICATION AS INFERRED FROM COMPLETE GENOME SEQUENCES

M. Naval-Sanchez, S. McWilliam, A. Reverter, M. Perez-Enciso, N. J. Hudson, J. Kijas

CSIRO Agriculture Flagship, 306 Carmody Road, St Lucia, QLD 4067, Australia

SUMMARY

The phenotypic diversity present within domestic sheep breeds is the outcome of direct human selection in behavioural as well as productive traits such as meat, milk or wool. To explore the genomic diversity of domestic sheep breeds we made use of whole-genome sequences of 68 domestic sheep sampled from five major geographic regions: Africa, the Americas, Asia, Europe and the Middle East. SNP calling identified a total of 26 million variants, ranging from 22 to 25 million SNPs per individual. The Asian and African animals examined contain a higher rate of heterozygosity (32.3% and 28.4%) compared to individuals from Europe (19.9%), the Americas (19.4%) or UK (20.4%). This is most likely a consequence of the sheep reference genome being from a European breed. In the future, we aim to compare these genomes against wild ovids to give further insight into the genomic mechanisms underlying domestication across breeds as well as their functional implications.

INTRODUCTION

Specific to sheep, the process of domestication was probably initiated around 11,000 years ago to facilitate stable access to meat and subsequently 5,000 years ago human mediated selection for wool and milk production (Chessa *et al.* 2009). Early consequences of animal domestication are likely to have included changes in stature, coat pigmentation, horn morphology in ruminants and docility (Zeder 2008). Genome-wide patterns of variation have proven highly informative for detecting genes under selection, with recent examples including loci controlling digestion (Axelsson *et al.* 2012), fertility (Larkin *et al.* 2012), stature and pigmentation (Rubin *et al.* 2010, Rubin *et al*. 2012) and horn development (Kijas *et al.* 2012).

In this preliminary analysis of 68 domestic sheep genomes, we compare patterns of genetic diversity and genetic divergence between individuals sampled from major geographic regions. This represents a first step in the reconstruction of the early evolutionary history of domestic sheep and the identification of loci involved in shaping the phenotypic diversity of today's modern breeds.

MATERIALS AND METHODS

Samples. Sixty-eight domestic sheep were sequenced using Illumina paired-end technology. Of these, 46 animals were selected from the ISGC Breed Diversity Hapmap experiment genotyped using the SNP50 Beadchip (Kijas et al. 2012), 6 animals were previously used for SNP discovery in the construction of the SNP50 BeadChip and CNV detection and the remaining animals werw investigated for the first time in this work. The selected animals belong to 42 different breeds drawn for Asia (n=12), Africa (n=6), the Middle East (n=13), the Americas (n=8), the United Kingdom (n=7) and continental Europe (n=22).

Alignment and variant calling. Reads from each sample were mapped against the sheep reference assembly v3.0 (available at http://www.livestockgenomics.csiro.au/sheep/) with BWA (Li and Durbin 2009) using default parameters. Duplicate removal and sorting were performed using samtools v.0.1.18 (Li et al. 2009). Genotypes were called for each animal separately using samtools mpileup. A series of filters were applied to prune low quality variants, including minimum depth of coverage (6 fold), map quality score (> 20) and base pair quality (>20).

Variants from each animal were then combined to produce a merged VCF file. This included examination to distinguish between positions with insufficient data to assign a genotype from those that were homozygous for the allele present in the reference genome.

Sequence based diversity estimates. To examine genomic differences among breeds and to infer population diversitys we made use of two metrics, namely principal components analysis (PCA) based on genetic diversity (heterozygosity level) and the compression efficiency (CE) algorithm (Hudson et al. 2014).

CE algorithm: In brief, CE is a new measure that exploits the order and proportion of heterozygosity in SNP genotypes. First, genotypes are encoded in numerical values 0's 1's or 2's for detected in bi-allelic SNPs across samples. Second, CE is calculated as CE= $(S_b\text{-}S_a) / S_b$, where S_b and S_a correspond to the size in bytes of the SNP genotype data before and after compression by the command gzip in UNIX, respectively. This measure is a proxy for the minimum amount of information required to reproduce a dataset. CE has shown to unravel genomic patterns such as phylogeography in diverse populations including human (Hudson et al. 2014).

Fixation Index (Fst): Fst to calculate the genetic distance between populations was calculated as in Weir and Cockerman 1984 paper, using vcftools *–weir-fst-pop* option.

RESULTS AND DISCUSSION

Whole genomes of 68 domestic breeds from different geographical regions Africa (n=6), Americas (n=8), Asia (n=12), Europe (n=22), Middle East (n=13), United Kingdom (n=7) were sequenced at an average depth of 8X in all groups (7.6-8.2) (Table1). SNP calling resulted in the discovery of a total of 26 million SNPs across the collection of animals. The average number of variants observed was calculated after grouping individuals into the geographically defined groups. The highest average number was identified in European animals, however this reflects the larger number of genomes sequenced. Next, we examined the percentage of heterozygous SNPs between populations and discovered that Asian and African populations contain a higher rate of heterozygosity (32.3 and 28.4) compared to breeds in Europe (19.9), Americas (19.4) and UK (20.4). This is most likely a consequence of the sheep reference genome reference being from a European breed (Jiang *et al.* 2014). Therefore, rather than considering it a measure of heterozygosity within breeds it reflects that Asian and African sheep are more genetically divergent to the reference genome in comparison to European, UK and American. Also, we calculated the Fixation Index (Fst) of each population compared to Middle East breeds, where first sheep domestication took place In all comparisons we observe very low Fst values showing a very weak population structure across sheep breeds (Table 1).

Table 1. Summary statistics on samples depth, number of called SNPs and percentage of heterozygosity

Region	Number of Samples	Average Depth	Average Number of called SNPs, millions	% Heterozygosity	Fst
Africa	6	8.2	24.34	28.3	0.024
Americas	8	8.1	23.18	19.4	0.021
Asia	12	8.2	23.86	32.3	0.020
Europe	22	8.1	25.05	19.9	0.023
Middle East	13	7.8	23.00	24.6	-
United Kingdom	7	7.6	22.23	20.4	0.030

Previous analysis based on mitochondrial haplotypes and SNP chip datasets have suggested globally distributed populations of sheep exhibit generally weak population substructure in comparison to other domestic species (Meadows et al. 2005; Kijas et al. 2012). We sought to determine if the much higher density (and unbiased) SNP collection obtained here is able to provide additional detail about the relatedness amongst a global collection of domestic sheep. We performed PCA of pairwise allele sharing to infer global patterns of genetic structure, with the results shown in Figure 1A. PC1 separated European and UK sheep from African, Asian, and Middle East. This largest PC only explained 4.2% of the total variance, indicating geographic origin is not a major source of variation. The second PC (2.6%) separated African, Middle East and Asian sheep. Finally, sheep from the Americas do not form a discrete cluster, but were rather distributed throughout the clusters ofAfrican or European animals. Thus, likely reflecting the highly admixed population history of the animals sampled from the Americas. Also, we analysed the 68 domestic sheep genotypes on basis of their compression efficiency (CE) and heterozygosity. The CE algorithm provides a new alternative to cluster populations based on the allele order and proportion across individuals (Hudson et al. 2014). It has been previously shown to reveal population structure in human populations, as well as cattle, mouse, dog and feral versus domestic sheep (Hudson et al. 2014). Here, we explore only domestic breeds which present relatively similar heterozygosity and CE levels (Figure 1B). Therefore, surprisingly, CE does not capture the same population structure as PCA and it is not able to clearly differentiate the phylogeography of the different breeds. Finally, the CE presents two clear outliers. The first corresponds to an Asian Garut animal with very low heterozygosity and high CE, whereas the second belongs to an American sheep from Santa Ines, with high heterozygosity and low CE. Possibly reflecting the level of admixture in different breeds.

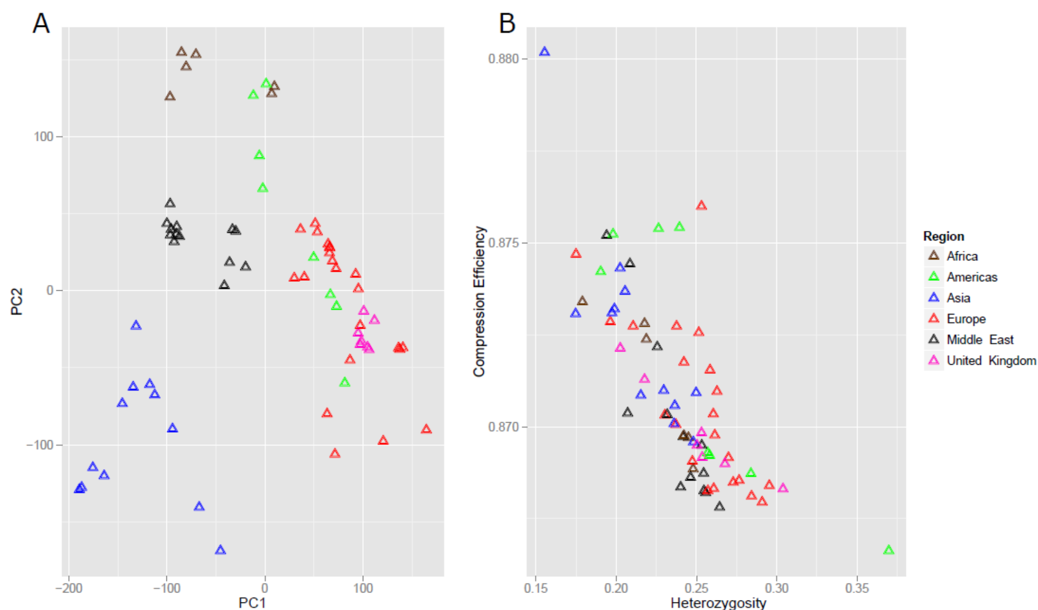


Figure1. Population structure. Breeds were coloured by origin Africa, Americas, Asia, Europe, Middle East and United Kingdom. A) Principal Component Analysis of genetic distance and B) plot CE versus Heterozigosity based on 550,048 SNPs without missing genotypes across the 68 animals.

Future prospects in our analysis is to study the genomic features selected in particular domestic breeds together with the addition of 18 wild ovid genotypes which would allow us to study the impact of domestication by defining genomic regions and the associated functional traits selected across domestic breeds.

REFERENCES

Axelsson E., Ratnakumar A., Arendt M-L., Maqbool K., Webster M.T., *et al*. (2013) *Nature* **495**: 360.

Chessa B., Pereira F., Arnaud F., Amorim A., Goyache F., *et al.* (2009) *Science* **324**: 532.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., *et al.* (2011). *Bioinforma Oxf Engl* **27**: 2156.

Gautier M.m Laloe D. and Moazami-Goudarzi K. (2010) *PlOS ONE* **5**:e13038

Hudson N.J., Porto-Neto L.R., Kijas J., McWilliam S., Taft R.J., *et al*. (2014) *BMC Bioinformatics* **15**: 66.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., *et al.* (2012) *PLoS Biol* **10**: e1001258.

Jiang Y., Xie M., Chen W., Tallbot R. Maddox J.F., *et al.* (2014) *Science* **6**:344(6188):1168.

Larkin D.M., Daetwyler H.D., Hernandez A.G., Wright C.L., Hetrick L.A., *et al.* (2012) *Proc Natl Acad Sci U S A* **109**: 7693.

Li H. and Durbin R. (2009). *Bioinforma Oxf Engl* **25**: 1754.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., 1000 Genome Project Data Processing Subgroup. (2009) *Bioinforma Oxf Engl* **25**: 2078.

Rubin C-J., Zody M.C., Eriksson J., Meadows J.R.S., Sherwood E., *et al.* (2010) *Nature* **464**: 587.

Rubin C-J., Megens H-J., Martinez Barrio A., Maqbool K., Sayyab S., *et al.* (2012) *Proc Natl Acad Sci U S A* **109**: 19529.

Tishkoff S.A., Reed F.A., Friedlaender F.R., Ehret C., Ranciaro A. , *et al*. (2009) *Science* **324**:528.

Weir B.S. and Cockerham C.C. (1984) *Evolution* **38**:1358.

Zeder M.A. (2008) *Proc Natl Acad Sci U S A* **105**: 11597.

ASSESSING IMPUTATION ACCURACY USING A 15K LOW DENSITY PANEL IN A MULTI-BREED NEW ZEALAND SHEEP POPULATION

R.V. Ventura[1,2], M. Lee[3], S.P. Miller[1,4], S.M. Clarke[4]and J.C. McEwan[4]

[1]Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada
[2] Beef Improvement Opportunities, Guelph, Ontario, Canada
[3] Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand
[4]Invermay Agricultural Centre, AgResearch Limited, Mosgiel, New Zealand

SUMMARY

Imputation has enabled genomic selection in commercial livestock, taking advantage of a more cost effective Low Density (LD) panel, increasing the number of genotyped animals and hence accelerating the adoption process. A 5K LD panel has been employed commercially in New Zealand. This study investigated the accuracy of imputation to 50K and High Density (HD) panels using a new 15K panelbeing developed by the International Sheep Genomics Consortium in four scenarios across two multi-breed New Zealand sheep populations. Theprototype panel resulted in higher values of imputation accuracy compared with the current LD panel (5K), which willbenefit the implementation of genomic selection for the sheep industry in New Zealand.

INTRODUCTION

Imputation is a robust tool able to infer the genotype at anun-genotyped locus and has been largely adopted for minimizing costs of genotyping in livestock breeding including sheep in New Zealand. Imputation assessment using the 5K LD panel in sheep was previously reported by Australian researchers (Hayes *et al.* 2012). In the New Zealand sheep industry, application of the current version of the low density panel (5K) has identified several genomic regions where imputation accuracy could be improved (Ventura et al. 2015, paper in submission), which could increase the accuracy of genomic predictions and further improve the identification of regions associated with traits of economic importance.A new 15K panel (in the process of design),containing markers selected by the International Sheep Genome Consortium (ISGC - www.sheephapmap.org),was used in this study to investigate imputation accuracies from 15K to both 50K and High Density (HD) panels in a multi-breed sheep population, pointing to potential regions for improvement over the 5K LD panel,which is used commercially for genomic selection in New Zealand sheep.

MATERIALS AND METHODS

Population imputation was implemented using the FIMPUTE 2.2 software(Sargolzaei *et al.* 2014). A total of 15,443 animals,part of the Beef and Lamb NZ genetics program (formally Ovita), composing a multi-breed sheep population, were genotyped with the Illumina OvineSNP50 Genotyping BeadChip(53,903 markers) and used in the present study to investigate the imputation from the low density panels (LD) 5K and 15K to the 50K panel. A second group of animals, part of the FarmIQ project, were genotyped using the Ovine Infinium® HD SNP BeadChip (606,006 markers) and were used to carry out the imputation from 15K to the HD panel. The HD animals were selected from eight flocks predominately of terminal composite breeds. Many of the animals were from recent breed developments (http://www.focusgenetics.com/) with undefined breed ratios and are best described as composites. The majority of the animals (~97%) in the second group were born in the period 2010 to 2013. The total number of animals used in the terminal composite population was 2,868, where 300 of these were used as a validation set and were born in 2013.The same strategy of using the youngest animals to be imputed was applied for the

population with 50K genotypes. Only autosomal markers were included in this investigation.For the imputation from LD to the 50K level, 12,853 markers (referred to 12K subsequently) out of the new 15K LD panel, remained after quality control.For the imputation from 15K to HD, 14,844 markers from the new LD set of SNPs were located on the HD panel and remained after quality control. Table 1 shows seven scenarios: six covering the imputation from 5K and 12K to the 50K panel in Romney, Coopworth and Perendale animals (Scenarios 1_R, 2_C and 3_P, respectively) and an additional scenario (4_TC) investigating the imputation from 15K to HD in the terminal composites. All LD panels used in this study were simulated by keeping markers in common between the respective LD and higher density panels and deleting remaining markers exclusively located in the higher density set (50K or HD).

Table 1. Description of imputation scenarios from 5K and 12K to 50K, and from 15K to HD panel

Scenarios	No. reference animals	No. imputed animals	Reference animals description	Imputed group breed[2]	Density
1_R	4256	1000	Romney	Romney	5K &12K to 50K
2_C	15443	250	All breeds	Coopworth	5K & 12K to 50K
3_P	15443	250	All breeds	Perendale	5K & 12K to 50K
4_TC	2568	300	Terminal composite breed	Terminal composite breed	15K to HD

Imputation accuracy was investigated using the allelic squared Pearson correlation (r^2) and concordance rate (CR), determined as the proportion of the correctly imputed markers out of all markers that were inferred after imputation. In both cases, the imputed and true genotypes (before deletion to build the LD panel) were compared. Common SNPs between LD and HD panels (15K) were not considered during the imputation accuracy determination.

RESULTS AND DISCUSSION

The accuracy of imputation from 5K to 50K ranged from 87.89% to 89.97% using the concordance rate measure and from 65.42% to 68.22% when the r^2 per SNP was calculated (Table 2). Concordance rate, calculated per animal or SNP, provides the same accuracy. Accuracies determined using r^2 per SNP marker,as done in this study,are usually lower than values calculated based on the animal, mainly due to the number of markers that are taken in consideration for the correlation estimates. An average gain in accuracy of 5.68% and 16.07% in CR and r^2, respectively, was noted after using the new 12K panel as the LD panel rather than the current 5K. The imputation from 15K to HD, performed in the second group of animals (terminal composite group), resulted in a CR imputation accuracy of almost 98% and squared correlation (r^2)of 88.70%.

Table2: Imputation accuracy under different scenarios

Scenario	5KCR	5Kr^2	12KCR	12Kr^2	15KHDCR	15KHDr^2
1_R	89.07	67.06	94.77	83.24	-	-
2_C	89.94	68.22	94.92	82.96	-	-
3_P	87.89	65.42	94.26	82.70	-	-
4_TC	89.03	-	-	-	97.81	88.70

The accuracy of imputation (CR) per animal, from 5K and 12K to 50K, is presented in Figure 1-left. Accuracies for almost all individuals were substantially increased by adding markers in the sparser panel and the largest gains in accuracy using the 12K panel (5K + new markers) were obtained for animals that obtained the lowest accuracies with the sparser panel (5K). Figure 1-right shows accuracy of imputation from 15K to HD where all individuals had their missing genotypes inferred with at least 90% success.

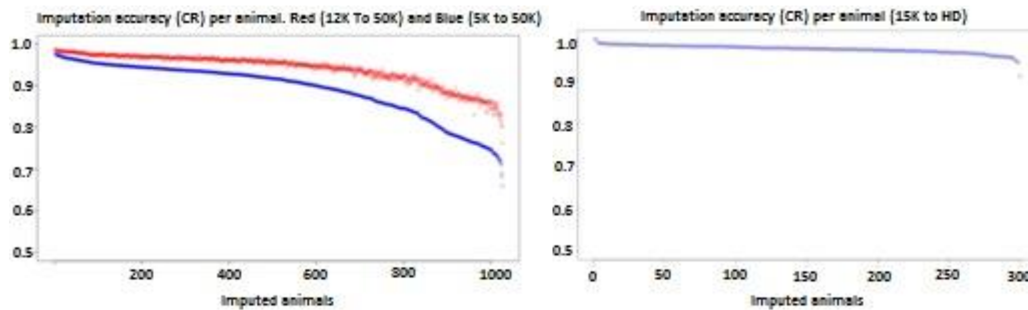


Figure 1. Accuracy of imputation per animal: Left – imputation from 5K(blue) and 12K(red) to 50K panel in Romney animals (Scenario 1_R). Right: imputation from 15K to HD panel in terminal composite breed. (In both plots X is reported as number of imputed animals and Y, as the CR measure of imputation accuracy ranging from 0.5 to 1.0).

A considerable increase in imputation accuracy to 50K for the 12K panel compared to the previous 5K panel was observed as illustrated in Figure 2. Almost all regions had higher imputation accuracy imputing from the 12K panel as compared to the 5K considering both r^2 and CR as metrics. The first 20Mb illustrates a region where the accuracy is improved considerably with the 12K panel. As illustrated in Figure 2, CR imputation accuracy per marker was higher than r^2 for all three scenarios. As reported by several authors in other species(Bouwman and Veerkamp 2014; Sargolzaei *et al*. 2014),imputation of markers at low MAF have lower r^2 accuracy than regions with higher MAF as can be noted by comparison of r^2 accuracy associated with MAF across different regions. The same pattern of increased imputation accuracy can be noted in the last two plots in Figure 2, where the increased accuracy showed almost the same trend for the same regions even in different populations and with higher density panels.

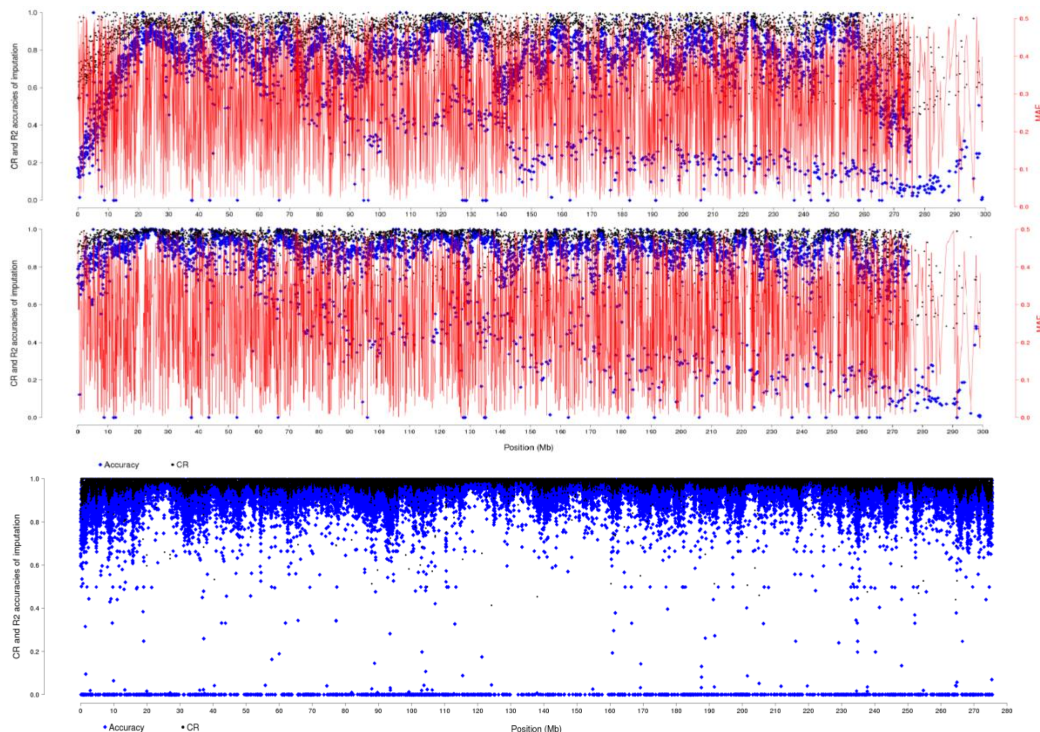


Figure 2 – Imputation accuracy per SNP evaluated by CR(black dots) and r^2(blue dots) according to the minor allele frequency (MAF is represented by the red line). First plot on top shows imputation accuracy from 5K to 50K for Chr1 and the central plot investigated imputation from 12K to 50K in Romney animals (Scenario 1_R). Last plot on the bottom shows imputation from 15K to HD in a terminal composite breed (Scenario 4_TC). (X is reported as Position (Mb) and Y, as the CR and r^2 measures of imputation accuracy ranging from 0.5 to 1.0)

The new 15K panel is still under development by the ISGC and the test results of this prototype panel will be used to inform the final panel implemented. Better imputation accuracy, especially at the chromosome ends can be expected with the new panel when implemented in New Zealand sheep.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the following organizations: Beef and Lamb New Zealand Genetics for access to 50K genotypes and FarmIQ (Ministry for Primary Industries' Primary Growth Partnership fund), for access to the HD genotypes, New Zealand sheep breeders for submitting samples, AgResearch Animal Genomics staff for genotyping and the ISGC for providing the prototype 15K panel design.

REFERENCES

Bouwman A. C. and Veerkamp R.F. (2014) *BMC Genet.* **15**:105.

Hayes B. J., BowmanP. J.,DaetwylerH. D.,Kijas J. W., and van der WerfJ.H.J.(2012) *Anim. Genet*. **43**:72.

Sargolzaei M., Chesnais J.P. and Schenkel F.S.(2014) *BMC Genomics* **15**:478.

ESTIMATING THE GENETIC (CO)VARIANCE EXPLAINED PER CHROMOSOME FOR TWO GROWTH TRAITS USING A HALF SIB DATA STRUCTURE IN SHEEP

C. Esquivelzeta-Rabell[1], N. Moghaddar[1,2], S. Clark[1,2] and J.H.J. van der Werf[1,2]

[1]School of Environmental and Rural Science, University of New England, NSW, Australia
[2]Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW, Australia

SUMMARY

To detect how much genetic variance is accounted for by different genomic regions one first step is to work at the chromosomal level. We used a half sib data structure for two growth traits in sheep as a potentially powerful design to partition the genetic variance across chromosomes. Records for post weaning weight (PW) and scan C site back fat (CF) were used from 5,239 merino sheep. The model of analysis accounted for population structure by fitting genetic group effects as well as the numerator relationship matrix (A) or the first five principal components (PC). Different approximations were compared fitting the genomic relationship matrix (G) based on 48,599 markers, or on single nucleotide polymorphisms of an individual chromosome. The correlation between chromosome length (L) and variance explained per chromosome ($\sigma^2_{g_i}$) was 0.53 and 0.70 for PW and CF correspondingly, however significant differences in ($\sigma^2_{g_i}$/L) were found between chromosomes, ranging from 0% to 17.5%. Some chromosomes explained more variance and covariance than expected, under the assumption that it is proportional to the chromosome size; suggesting that some chromosomes clearly harbor more QTL. Some chromosomes show a covariance of opposite sign indicating they could be used in selection to 'break' an unfavourable correlation (e.g. chromosome 8). These results represent a powerful source of information for genomic selection.

INTRODUCTION

The tracking of chromosome segments through a pedigree is becoming more feasible due to the availability of abundant genomic information. However with the ever increasing density of genetic markers, there is also an increasing ambition to work out which variants actually are responsible for the observed quantitative genetic variation.

One first step to detect how much genetic variance is accounted for by different genomic regions is to work at the chromosomal level. Previous studies found a linear relationship between chromosome length and variance explained for human traits (Yang *et al.* 2010 and Visscher *et al.* 2007), production and fitness related traits in dairy cattle (Jensen *et al.* 2012) and production traits in sheep (Daetwyler *et al.* 2012).

The advantage of data on sheep populations is that the data structure is usually based on relatively large half sib families. This provides a powerful design for determining segregation based on linkage. The design is not suitable for LD mapping; hence the accuracy of mapping QTL positions would be low. However, the latter is less relevant for determining the amount of genetic variance explained per chromosome.

The objectives of this study were to estimate the genetic variance and covariance for two growth traits in sheep, determine the amount of additive genetic variance explained by each chromosome and to investigate the best model to correct for population and pedigree structure.

MATERIAL AND METHODS

Data for this study was obtained from the Information Nucleus program of the CRC for Sheep Industry Innovation. Details on this program and its design are described by van der Werf *et al.*

(2010). The data set comprised a total of 2,455 purebred merino lambs with phenotypes for two growth traits (post weaning weight: PW and scanned C site back fat: CF), pedigree and genotype data. The animals were descended from 139 sires and the associated pedigree file contained 10,559 animal identities from over 22 generations. The pedigree information was used to compute a numerator relationships matrix (A) for the animals with phenotypic records using the R package 'pedigree' (Coster 2012). Genotypic information consisted of SNP marker genotypes obtained using the OvineSNP50 BeadChip assay (Illumina, San Diego, USA). After quality control (Moghaddar *et al.* 2014) and imputing missing genotypes with Beagle 3.2 (Browning and Browning 2007), genotype information on 48,599 SNP was used to derive a genomic relationships matrix (G), scaling G to be analogous to A following VanRaden (2008).

The general model used to analyse the data was: $y = Xb + Z_a a + Z_m m + Z_q Qq + e$, where vector b included fixed effects of sex of lamb (ram: 1 or ewe: 2), birth type/rearing type (single: 1/1, twins: 2/2 or triplets: 3/3 and their combinations), management group, age of dam and weaning age; a is the random additive genetic effect of the lamb, m is the maternal permanent environmental effect and q is a genetic group effect. The genetic group consisted of merino strain (depending on the type of wool) where we regressed on strain proportion. Different models were explored to be able to partition the additive genetic variance into components that can be explained by markers while correcting for population structure, fitting into the model A and G individually (model 1 and 2 correspondingly) or simultaneously (model 3), together with genetic groups (merino strain) derived from a deep pedigree analysis, or fitting the first five principal components (PC) as a covariate (model 4). To decompose the variance components into 26 chromosomes, individual G_i were built based on marker information on the ith chromosome and fitted simultaneously in the model (model 5), together with A (model 6) and with PC as a covariate (model 7). A bivariate analysis was also performed using the estimated variance components per trait resulting from the univariate analysis to define the starting values of the (co)variance matrices structures. The variance components for the first 4 models and the bivariate analysis were analysed using ASReml 3.0 software (Gilmour *et al.*, 2009). Models 5 to 7 were analysed using GCTA software (Yang *et al.*, 2011b).

RESULTS AND DISCUSSION

When pedigree (**A**) and marker based (**G**) relationship matrices were fitted individually we found that for PW SNPs capture additive genetic relationships among individuals and also effects of QTL (Table 1), G explaining more variation (36.80 %) than A (30.81%); however in the case of CF the results were opposite, A explained slightly more variation (25%) than G (22.61%). Nevertheless for both traits the log likelihood was higher when fitting only G in the model compared to fitting only A. Results from model 3, in which A and G were fitted simultaneously showed that most of the variance was partitioned toward G agreeing with previous reports (Haile-Mariam *et al.* 2013 and Jensen *et al.* 2012).

Model 4 was investigated as an alternative solution to correct for population structure avoiding co-linearity between the variance components. Results showed that variance explained by G after accounting for population structure using the first five PC was equal to 6.69 for PW and 0.09 for CF and the phenotypic variance was similar to the one calculated with the rest of the models.

Genetic variance explained per chromosome using models 5 to 7 showed that the variance explained by each G_i varied and was somewhat related to the chromosome length (Figure 1) agreeing with Daetwyler *et al.* (2012). The correlation between variance explained and chromosome length was weaker for PW (0.55) than for CF (0.70) and marked differences in genomic variance explained were found for some chromosomes. For example, chromosome 6 explaining the higher amount of genomic variance for PW (13.3 to 13.7% depending on the model), followed by chromosome 1 (8.6 to 8.8%), 2 (6.4 to 8.2%), 10 (7.6 to 8.9%), 16 (5.8 to

6.4%) and chromosome 26 is estimated to contribute 0% variance.

Table 1. Variance components estimates using different mixed linear models for post weaning weight (PW) and scanned C site back fat (CF).

PW	σ_a^2	σ_g^2	σ_m^2	σ_e^2	LogL
Model 1 (A)	6.74		2.21	12.92	-9485.64
Model 2 (G)		7.75	1.77	11.54	-9393.23
Model 3 (A+G)	1.33	7.74	1.44	10.88	-9393.19
Model 4 (G+PC)		6.69	2.38	12.12	-10358.2
CF					
Model 1 (A)	0.12			0.36	-1293.60
Model 2 (G)		0.11		0.37	-1279.29
Model 3 (A+G)	0.03	0.10		0.35	-1279.09
Model 4 (G+PC)		0.09		0.43	-1197.29

The difference between variance explained per chromosome and the expected variance explained, under the assumption that the genetic variance is proportional to the size of the chromosome (Daetwyler *et al.*, 2012), was also calculated (Figure 1) showing that some chromosomes explain more variance than expected; reflecting the relative QTL density on each chromosome, e.g. chromosome 6 for both traits with and extra variance of 9.1% for PW and 7.1% for CF; and others contribute with 0% of the variance explained, e.g. chromosomes 8, 9, 14, 19 and 25 for CF.

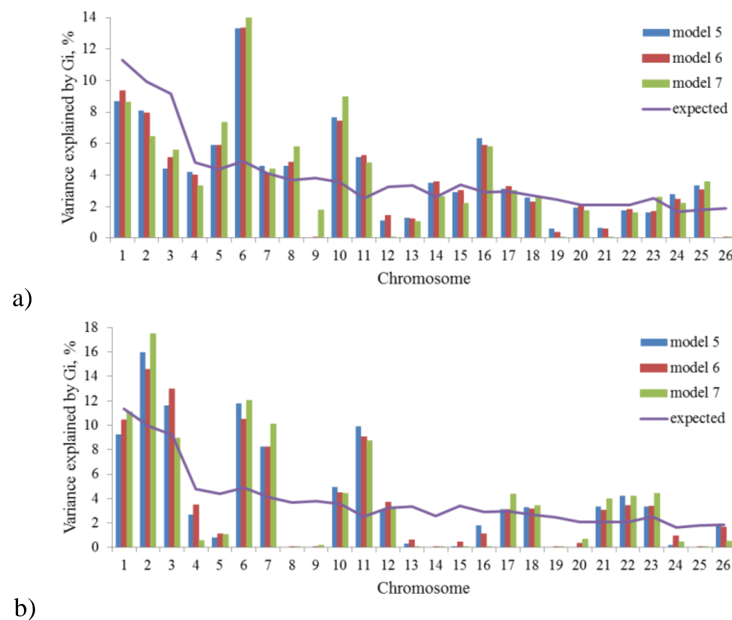


Figure 1. Genomic proportion of variance relative to the total size of the genome calculated per chromosome for post weaning weight (a) and scanned C site back fat (b) using models 5, 6 and 7. The line indicates the expected proportion of genomic variance explained per chromosome based on its size.

A bivariate model was used to estimate covariances and correlations per chromosome between growth traits in sheep. We found that 6 chromosomes have large covariance effects (Figure 2) and 5 have a negative covariance, e.g. chromosome 8, revealing that a small number of chromosomes appeared to 'break' the positive genetic correlation (overall genetic correlation ~ 0.5).

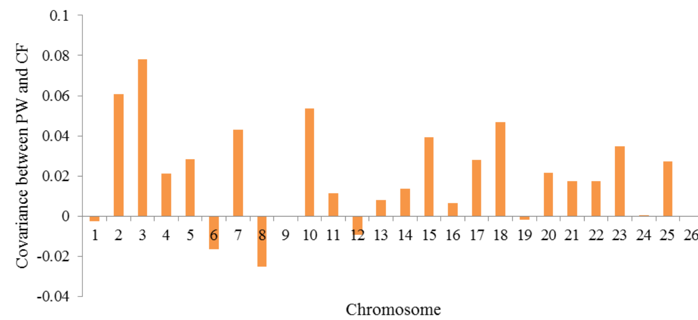


Figure 2. Covariance between post weaning weight (PW) and scanned C site back fat (CF) in merino lambs from bivariate analysis per chromosome.

The main conclusions from the present study are that the inclusion of PC in the model corrects for population structure avoiding co-linearity between the variance components. While the additive genetic variance explained per chromosome is partially related to chromosome length, considerable differences between chromosomes in the amount of additive genetic variance explained were found and a small number of chromosomes appeared to 'break' the positive genetic correlation.

The approach presented in our study provides relevant information to the understanding of the genetic underlying complex trait variation and represents a powerful source of information for genomic selection.

REFERENCES

Browning S.R. and Browning B.L. (2007) *Am. J. Hum. Genet.* **81**: 1084.

Coster A. (2012). pedigree: Pedigree functions. R package version 1.4. http://CRAN.R-project.org/package=pedigree

Daetwyler H.D., Kemper K.E., van der Werf J.H.J. and Hayes B.J. (2012) *J. Anim. Sci*. **90**: 3375.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Haile-Mariam M., Nieuwhof G.J., Beard K.T., Konstatinov K.V. and Hayes B.J. (2013) *J. Anim. Breed. Genet*. **130**: 20.

Jensen J., Guosheng S. and Per M. (2012) *BMC Genetics*. **13**: 44.

Moghaddar N., Swan A.A., van der Werf J.H.J. (2014) *GSE.* **46**: 58.

van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Anim. Prod. Sci*. **50**:998.

VanRaden P.M. (2008) *J. Dairy Sci*., **91**: 4414.

Visscher P.M., Macgregor S., Benyamin B., Zhu G., Gordon S. *et al.* (2007) *Am. J. Hum. Genet*. **81**: 1104.

Yang J., Benyamin B., McEvoy B., Gordon S., Henders A.K. *et al.* (2010) *Nat. Genet*. **42**: 565.

Yang J, Lee SH, Goddard ME and Visscher PM. (2011) *Am. J. Hum. Genet*. **88**(1): 76.

RUNS OF HOMOZYGOSITY IN SWAKARA PELT PRODUCING SHEEP: IMPLICATIONS ON SUB-VITAL PERFORMANCE

F.C. Muchadeyi[1], M.T. Malesa[1,2], P. Soma[3] and E.F. Dzomba[2]
[1]Agricultural Research Council, Biotechnology Platform, Onderstepoort, 0110, South Africa
[2]Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Scottsville 3209, South Africa
[3]Agricultural Research Council, Animal Production Institute, Irene 0062, South Africa

ABSTRACT

Sub-vital performance is a phenotype in which some pure white Swakara sheep die within 48 hours of birth. Postmortem of sub-vital lambs have shown underdeveloped digestive organs. It is hypothesized that due to high levels of inbreeding, Swakara sheep carry a recessive mutation that affects some of the white fleece colour subpopulation resulting in the sub-vital production performance. The genetic basis of the sub-vital effect is however unknown. The aim of this paper was to use the Ovine SNP50K data to investigate inbreeding levels and occurrence of ROH in the Swakara sheep genome in order to uncover the genetic basis of sub-vital performance. Runs of homozygosity (ROH) are long stretches of contiguous DNA fragments that are homozygous and occur due to parents transmitting similar haplotypes to their offspring through shared ancestry. Long stretches of ROH observed in SNP markers indicate probable underlying stretches of DNA inherited identical by descent from recent common ancestors. The aim of the study was to screen for extended ROH shared across animals, estimate levels of inbreeding using ROH and infer association with sub-vital performance which could indicate regions associated with the sub-vital mutation. Ninety-four Swakara individuals that belonged to the grey (n =22), black (n=15) and white vital (n = 41) and white sub-vital (n=16) were genotyped using the OvineSNP50 beadchip. Four hundred and thirty six unique ROH regions that spanned between 1001 to 6594 Kb were observed on the 25 chromosomes in 94 individuals of the four colour subpopulations. Three consensus ROH (cROH) were more prevalent in sub-vital Swakara sheep. Results suggested alternative genetic mechanisms to sub-vital performance other than was initially hypothesised that sub-vital performance was due to recessive mutations prevalent in inbred white Swakara sheep.

INTRODUCTION

Swakara is a fat-tailed sheep breed that was improved through generations of selective breeding from the Karakul breed which originates from the Middle East (Campbell 2007; Soma *et al.* 2012). White pelts are preferred to other colours (brown, black and grey) on the market (Campbell 2007) because they can be dyed to any desired colour to make coats and other fashion products. Production of white pelt is however hampered by a sub-vital factor that affects some of the pure white Swakara sheep that die within 48 hours of birth. Postmortem of affected lambs have shown underdeveloped digestive organs. The pure white animals require special care that includes specialized diets and most die before reproductive maturity, even under special care. As a practice the breeders mate A-white (sub-vital pure white fleece) rams that are raised under specialized care with black ewes to produce 100% B-white and C-white progeny, which are predominantly white with black patches (and are vital). Mating of B-white X B-white (C-white) produces 25% black; 50 % B/C-white and 25% A-white (sub-vital).

It is hypothesized from the products of the different matings that the appearance of the sub-vital factor in pure white Swakara could have resulted from intensive selection leading to the accumulation of the homozygous recessive allele associated with sub-vital performance. However, there are other white Swakara sheep that do not experience sub-vital performance implying some level of independent segregation of the sub-vital loci from the coat colour loci in some

populations. An understanding of the genomic structure and diversity of the white vital, white sub-vital and the two other Swakara colour variants (black and grey) is important in determining the genetics of sub-vital performance and to aid breed improvement programmes.

Runs of homozygosity (ROH) are defined as long and continuous stretches of homozygous genotypes (Ku *et al.* 2011) that result from parents transmitting identical haplotypes to their offspring. ROH have been used as an estimate of inbreeding (Zhang *et al.* 2013). Recent studies have found associations between regions of extended homozygosity with both complex and simple gene disorders (Kijas *et al.* 2012; Suarez-Vega *et al.* 2013). Due to the small population sizes of Swakara breeding flocks and the intense selection, high levels of inbreeding and longer extended ROH could be occurring at a high frequency and also contributing to the expression of the sub-vital factor. In this study, we investigated inbreeding levels and screened the genome of Swakara sheep for the occurrence of ROH using the Ovine SNP50K data. The genes covered by ROH were investigated and inference made on associations with sub-vital performance.

MATERIALS AND METHODS

Animal genotyping and quality control. Blood was collected from 96 Swakara sheep sampled from two research stations and 3 farms in Namibia (n=60), Carnavon Research Station in South Africa (n=30) and consisting of four colour sub-populations i.e. black (n=16), grey (n=22), white vital (n=35) and white sub-vital (n=17). The white-vital referred to a sub-population of white Swakara that did not experience sub-vital performance whilst the white sub-vital would die prematurely. The ovineSNP50 beadchip was used for genotyping. Quality control was done to remove SNPs where the genotype call rate was < 0.05, missing genotypes were >0.10, the minor allele frequency was < 0.05 or there was a deviation from Hardy Weinberg equilibrium ($P < 0.001$). At least 80.35 % of the SNPs were left for further analysis after quality control. Two individuals, one white sub-vital and one black, were excluded from downstream analysis due to missing genotypes.

Inbreeding and runs of homozygosity. Inbreeding coefficients (F_{IS}) and runs of homozygosity (ROH) were estimated from SNP genotypes using Plink v.1.7 (Purcell *et al.* 2007). In this analysis, the ROH were defined as homozygous stretches along a 1000kb sliding window with a minimum of 20 SNPs allowed in the tract, with allowance of at most 2 missing SNPs and no heterozygous SNPs. The threshold for SNP density was set at 50kb/SNP. When two SNPs within a segment were too far apart (>500 kb), the segment was split in two using the --*homozygous gap* function. Pools of overlapping segments were estimated using PLINK. The size of the consensus region was determined as well as the distribution of the consensus region in the sub-vital and normal Swakara sheep.

Identifying potential trait association of SNPs in ROH segments. The Golden Helix GenomeBrowse® was used to identify genes within the consensus regions of ROHs using the sheep reference genome *Ovies Aries OAR v.3.1* and *Ensembl* (www.ensembl.org/sheep). The genomic base location of the first and last SNP of a ROH was used to denote the region of interest.

RESULTS AND DISCUSSION

Inbreeding levels and ROH patterns in Swakara subpopulations. The inbreeding levels of the four Swakara subpopulations are shown in Table 1. The Grey subpopulation was the least inbred with an F_{IS} value of 0.01 whilst the black Swakara had the highest F_{IS} of 0.09. The low inbreeding levels observed in the grey subpopulation were expected since homozygous grey karakul also experience a lethal factor documented by Groenewald (1993). The black sub-population is known not to experience any sub-vitality or lethal factors and is currently used in crossbreeding with pure white Swakara as a way of managing sub-vital

performance. Four hundred and thirty-six (436) ROH were observed in 94 Swakara sheep. All sheep had at least one reported ROH (Table 1), and there were an average of about 5 ROH events reported per animal in each subpopulation.

Chromosomal coverage of ROH in Swakara population. The ROH length ranged from 1001Kb to 6594 Kb and covered at least 13.96% of the Swakara sheep genome. There were no ROHs observed on chromosome 23. The highest chromosomal coverage was observed on chromosome 9 that had 35 unique ROHs covering 44% of the chromosome.

Table 1. Inbreeding (F_{IS}) levels and the characteristics and number of ROH Swakara sheep.

Subpopulation	F_{IS}	No. ind. with ROH	No. ROH	Mean ± SD ROH length (KB)	Mean ± SD NSNPs
White-vital	0.089±0.063	41	214	1661.13±788.92	35.01±16.07
White subvital	0.078±0.13	16	84	1610.18±804.56	34.19±16.76
Black	0.094±0.075	15	72	1567.67±776.72	33.28±15.32
Grey	0.011 ±0.069	22	109	1749.98±922.79	37.05±18.58
Overall	0.089±0.063				

Consensus (overlapping) ROH and putative function. The 436 unique ROH grouped into 131 consensual ROH (cROH). Of these only three on chromosomes 3, 4 and 25 had a distribution biased towards sub-vital Swakara sheep (Table 2). The consensus ROH (cROH) on chromosome 4 was observed in 10 sheep six of which were from the sub-vital group. Those cROH on chromosomes 3 and 25 were observed in 2 individuals each both of which were of sub-vital performance.

Table 2. The consensus ROH common in white sub-vital sheep and the associated genes.

cROH	Chr	No. of individuals	Length of cROH (Kb)	No. Sub-vital sheep	Genes within region
cROH1	3	2	1201.90	2	C2orf74
cROH2	4	10	2768.36	6	DPP6;HTR5A-AS1;SHH
cROH3	25	2	1018.51	2	LRRTM3

Consensus regions cROH1 and cROH3 on chromosome 3 and 25, respectively, were observed in only in 2 white sub-vital Swakara sheep whereas cROH2 on chromosome 4 was observed in 6. Using the GenomeBrowse® tool we found that cROH3 is within a genomic region carrying the leucine-rich repeat transmembrane 3 (LRRTM3) gene. This gene family is associated with late onset of Alzheimer's disease in mice (Laakso *et al*., 2012). The consensus region cROH2 on chromosome 4 is a 2.8MB region on which lies the DPP6, HTR5A-ASI and SHH genes. No information was found on the putative functions of HTR5A-ASI, while studies on DPP6 suggest it functions on the development of the brain and skeletal system in livestock (Buzanskas *et al*., 2014). The mammalian SHH has been reported to encode a signaling molecule that is vital for the developmental patterning especially of the nervous system and the skeletal system (Dorus *et al*., 2006). The trend in the observed cROH in white sub-vital Swakara sheep observed in this study suggests some candidate genes impacting on the nervous function and skeletal and brain development. The cROH2 carrying these genes was however also observed in a few (40%) of vital Swakara sheep which makes it a less likely candidate for sub-vital performance under the assumption that the sub-vital performance was a recessive mutation only experienced in white Swakara sheep. These results suggest alternative genetic mechanisms to sub-vital performance

other than was initially hypothesised that sub-vital performance was due to recessive mutations prevalent in inbred white Swakara sheep. Small sample sizes of both the vital and sub-vital sheep were a major limitation of this study. Alternative analyses such as genome-wide association studies and population stratification based analysis could be used to further investigate potential genes and the genetic mechanisms for sub-vital performance in white Swakara sheep.

CONCLUSION

Apart from the grey subpopulation, the Swakara sheep are moderately inbred with an average inbreeding co-efficient (F_{IS}) of about 10% percent. Our results do not support the presence of a recent recessive-lethal mutation causing the sub-vital phenotype in white Swakara sheep.

REFERENCES

Buzanskas M. E. Grossi D.A., Ventura R.V., Schenkel F.S., Sargolzaei M. *et al.* (2014) *PLoS ONE* **9**: 4.
Campbell L.J. (2007) MSc Dissertation. University of Pretoria, South Africa.
Dorus S., Anderson J.R., Vallender E.J., Gilbert S.L., Zhang L., *et al.* (2006) *Hum. Mol. Gen.* **15**: 2031.
Groenewald H.B. (1993) *Onderstepoort Journal of Veterinary Research* **60**: 189.
Kijas J.W., Lenstra J.A., Hayes B.J., Boitard S. *et al.* (2012) *PLoS Biology* **10**.e1001258.
Ku C.S., Naidoo N., Teo S.M. and Pawitan Y. (2011) *Hum. Genet.* **129:** 1.
Laakso T., Muggalla P., Kysenius K., Laurén J., Paatero A., *et al.* (2012) *J. Alzheimers Dis.* **4**: 759.
Purcell S., Neale, B., Todd-Brown K., Thomas L., Ferreira M., *et al.* (2007) *Am. J. Hum. Genet.* 81:559.
Soma P., Kotze A., Grobler J.P. and van Wyk J.B. (2012) *Small Rumin. Res.* **103:** 112.
Suarez-Vega A., Gutierrez-Gill B., Cuchillo-Ibanez I., Saez-Valero J., Perez V., *et al.* (2013) PLoS *One* **8**.e81072
Zhang H., Li J., Zhao F., Ren H., Xu L., *et al.* (2013) *PLoS ONE* **8**.e66569.

GENETIC DIVERSITY AND EFFECTIVE POPULATION SIZE OF EIGHT IRANIAN CATTLE BREEDS

E.M. Strucken[1], K. Karimi[2], A.K. Esmailizadeh[2], N. Moghaddar[1] and C. Gondro[1]

[1] School of Environmental and Rural Science, University of New England, Armidale 2351, NSW, Australia

[2] Department of Animal Science, Shahid Bahonar University of Kerman, Kerman, Iran

SUMMARY

Iranian cattle breeds are currently under-represented in studies regarding genetic variability and conservation effort even though they are settled in a region that is believed to be one of two ancient cattle domestication centres. This study provides first population genetic parameters for eight Iranian cattle breeds collected from across the entire country to highlight the necessity for conservation programs. High density genome-wide SNP chips were used. None of the Iranian breeds showed a decreased heterozygosity compared to outgroup breeds (Holstein, Jersey, and Brahman) and inbreeding coefficients were low. Nevertheless, estimated effective population sizes were <10 for the Mazandarani, Sarabi, and Kermani breeds and it is predicted that most of the genetic variability will be lost within 20-30 generations if no intervention measures are taken. Effective population size estimates varied between chromosomes with occasional extremely high values, especially for Najdi, Pars, and Kermani which have high proportions of indicine ancestry as represented by the Brahman outgroup.

INTRODUCTION

Two separate domestication events gave rise to the variety of cattle breeds we see today. India is the origin of humped zebu cattle, and the Fertile Crescent of the Near East is the region of origin for humpless taurine cattle (Loftus *et al.* 1994). Iran covers the Eastern side of the Fertile Crescent and the native cattle breeds represent an extensive biological resource for origin and domestication studies. Iran is home to a large number of cattle breeds, however, the number of indigenous animals is declining with breeds, such as the Golpayegani, becoming extinct and other indigenous breeds endangered. The loss of these breeds or their genetic diversity, which is the ultimate source of adaptive variation to environmental pressures, will significantly limit the genetic resources available to future breeding programs (Herrero-Medrano *et al.* 2013).

Genetic characterization of the Iranian breeds provides the first step towards breed conservation, which largely depends on existing genetic variability and effective population sizes. A few studies attempted to examine the genetic diversity of Iranian native cattle (Mirhoseinie *et al.* 2005; Valizadeh *et al.* 2012). However, these studies concentrated on only a few breeds and were based on small numbers of markers (Mirhoseinie *et al.* 2005; Valizadeh *et al.* 2012). In this study, a dense SNP dataset was used to investigate genetic diversity, inbreeding, and effective population sizes in eight Iranian native cattle breeds.

MATERIALS AND METHODS

Data. Hair samples of individuals from eight Iranian cattle breeds were collected throughout the country. The Iranian breeds could be clustered into four taurine populations (humpless), and four indicine populations expressing a hump and often a pronounced dewlap and pendulous ears. Additionally, randomly selected subsets from larger populations of Holstein, Jersey and Brahman cattle were included to anchor the Iranian breeds towards taurine and indicine origins.

All animals were genotyped with the 700k Bovine BeadChip (Illumina Inc, San Diego, CA, USA). Genotypes were subject to a stringent quality control (Gondro *et al.* 2014) with a GC score

>0.9, call rates per marker >90% and per animal >70% (reduced due to small sample size). Markers were excluded that deviated (1) in their heterozygosity by more than 3 standard deviations from the mean heterozygosity or (2) from Hardy-Weinberg equilibrium at P-value<10^{-16}. Only markers that were present in all breeds including outgroups were kept for further analyses which reduced numbers considerably (283,028 SNPs).

Analyses. Population parameters such as average minor allele frequency (MAF) and average heterozygosity (He) were calculated per breed. Inbreeding coefficients (F_{IS}) were estimated according to Weir and Cockerham (1984). Based on an unsupervised analysis in ADMIXTURE 1.23 (Alexander *et al.* 2009) for two ancestral populations we calculated the percentage of indicine origin (as represented by the Brahman population) in contrast to the taurine origin (as represented by the Holstein and Jersey populations) in the Iranian breeds.

Effective population sizes were estimated with the *LDNe* program (Waples and Do 2008). *LDNe* conducts population size estimations based linkage disequilibrium as represented by pairwise correlations (r^2) according to Hill (1981) and Waples (2006). Only markers with allele frequencies >0.01 were used and the mating system was assumed to be random. Estimates were carried out per chromosome. For the purpose of conservation genetics, changes in heterozygosity over time (generations 1-50) were estimated as $H_t = H_0 (1 - 1/(2 N_e))^t$ where H_t and H_0 are heterozygosity at generation t and generation zero, respectively, and N_e is the effective population size as estimated with the *LDNe* program and averaged across all chromosomes.

RESULTS AND DISCUSSION

The Iranian cattle breeds showed similar heterozygosity levels compared to the applied outgroup breeds. Inbreeding coefficients were close to zero indicating no apparent loss of genetic diversity (Table 1). The Pars breed had the highest inbreeding coefficient (0.121; Table 1). Surprisingly, the Holstein cattle showed an increase in heterozygosity as indicated by a negative inbreeding coefficient (-0.109; Table 1). This unexpected result indicates that further interpretation of results concerning the Holstein breed have to be taken in the context of this study, as Holsteins are generally reported to have a decreased heterozygosity due to strong artificial selection (Rodriguez-Ramilo *et al.* 2015).

Table 1. Breed description and population genetics parameters for 11 cattle breeds after quality control

Breed	N	N_e ±se	Appearance	MAF	He	F_{IS}	% indicine
Jersey	15	10.6 ±1.00	taurine	0.22	0.32	-0.041	0.00001
Holstein	15	7.7 ±0.91	taurine	0.24	0.36	-0.109	2.5
Kurdi	7	31.1 ±10.6	taurine	0.26	0.36	0.005	32.3
Sarabi	19	6.1 ±0.84	taurine	0.25	0.34	-0.023	42.1
Pars	7	63.7 ±28.4	taurine	0.21	0.26	0.121	80.9
Kermani	9	8.8 ±2.03	taurine	0.20	0.27	0.083	82.6
Taleshi	7	72.0 ±20.1	indicine	0.23	0.32	0.017	61.5
Mazandarani	10	3.1 ±0.87	indicine	0.23	0.32	0.018	65.6
Najdi	7	165.5 ±56.8	indicine	0.21	0.31	-0.001	75.1
Sistani	9	45.2 ±19.8	indicine	0.16	0.23	0.059	95.3
Brahman	15	35.0 ±5.99	indicine	0.16	0.23	0.004	99.9

Ne: effective population size; MAF: minor allele frequency; He: heterozygosity; F_{IS}: inbreeding coefficient

Notably, the more indicine a breed was (as represented by the percentage of Brahman origin) the lower the heterozygosity (Table 1). However, this trend can be explained by ascertainment bias

where indicine loci are often less variable due to mostly taurine breeds used in design of the SNP chip (Lachance and Tishkoff 2013). Interestingly, some of the breeds that were classified into taurine and indicine based on their outer appearance were re-grouped based on their genetic constitution. As such, Taleshi and Mazandarani expressed dewlaps and slight humps but had less indicine ancestry than Pars and Kermani (taurine appearance; Table 1). This discrepancy shows how important genetic analyses are to correctly assign individuals to breeds and ancestries.

Effective population sizes based on LD were on average lowest for Mazanderani, Sarabi, and Kermani (Table 1). For these breeds, a close monitoring of potential inbreeding depressions is warranted; however, the small sampling size might have led to a downward bias of estimates. The Sarabi breed is currently the most used dairy breed in the North-West of Iran and the small effective population size of this breed might also stem from a semi-established breeding program by the Iranian government. Thus, individuals for mating could be artificially restricted. Effective population sizes for Holstein and Jersey were much lower compared to previous reports (Rodriguez-Ramilo *et al.* 2015; Stachowicz *et al.* 2011) and the downward bias of our study is most likely due to the small sample size. Nevertheless, comparisons can be made within the results of this study.

Effective population sizes showed a large variation depending on the chromosome that was used for estimation. Whilst estimates of most chromosomes resulted in N_e <100, chromosomes 3, 6, 10, 11, 13, 20, 23, 25, and 29 showed highly increased effective population sizes (Figure 1). Breeds with chromosome-wise outstanding N_e were Najdi, Pars and Sistani which are all breeds with an increased indicine/Brahman proportion. Taleshi showed highest chromosome-wise N_e for the taurine classified breeds (Figure 1).

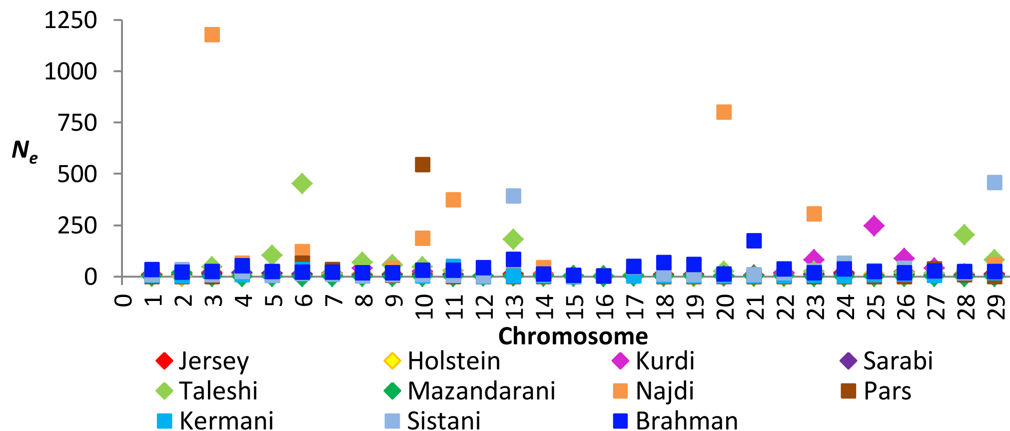


Figure 1. Effective population sizes (N_e) per chromosome for 11 cattle breeds.
diamonds: taurine breeds; squares: indicine breeds

According to Waples (2006; *equation 7*), N_e is dependent on LD (r^2) and sample size. Even though sample sizes in this study varied slightly between populations, we can deduce that chromosomes and breeds with outstandingly high N_e must have a very weak LD (small r^2). Linkage disequilibrium between markers is also an important measure to estimate the phenotypic variance that is explained by the genetic sequence. Esquivelzeta-Rabell *et al.* (2014) demonstrated in sheep that the variance explained differed between chromosomes. Even though not completely transferable between sheep and cattle, chromosomes with the highest explained variance were also the chromosomes in this study that show outstandingly high N_e estimates (except chromosomes

3 and 29).

Based on average heterozygosities and N_e estimates, genetic variability over the subsequent generations can be predicted if no intervention is made via conservation programs. Breeds with the smallest effective population sizes will lose almost all of their genetic variability within the next 20-30 generations (**Figure 2**). As previously stated, conservation programs should aim to maintain the genetic variability of these breeds to preserve this valuable resource for the future.

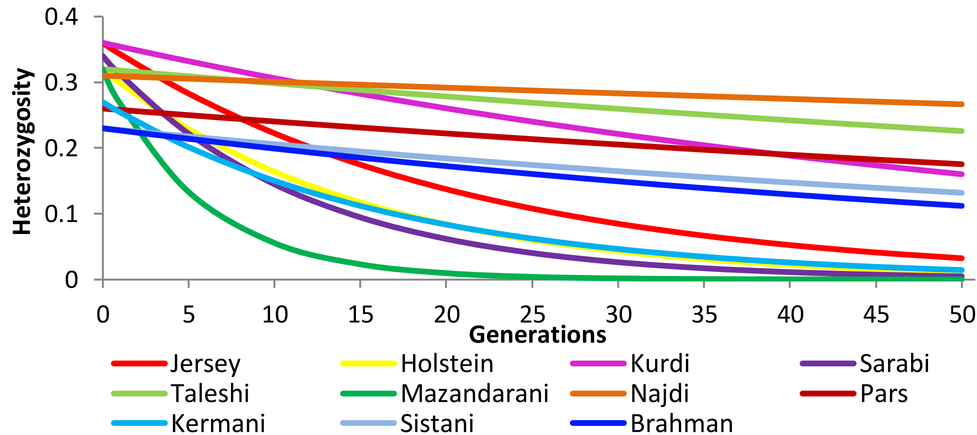


Figure 2. Decay of heterozygosity over 50 generations in 11 cattle breeds.

ACKNOWLEDGEMENTS

EMS and CG were supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ01134906), Rural Development Administration, Republic of Korea.

REFERENCES

Alexander D.H., Novembre J. and Lange K. (2009) *Genome Res.* **19:** 1655.

Esquivelzeta-Rabell C., Moghaddar N., Clark S. and Van der Werf J. (2014) *World Congress on Genetics Applied to Livestock Production*, Vancouver, CA.

Gondro C., Porto-Neto L.R. and Lee S.H. (2014) *Anim. Genet.* **45:** 758.

Herrero-Medrano J.M., Megens H.J., Groenen M.A., Ramis G., Bosse M., Perez-Enciso M. and Crooijmans R.P. (2013) *BMC Genet.* **14:** 106.

Hill W.G. (1981) *Genetics Res.* **38:** 209.

Lachance J. and Tishkoff S.A. (2013) *Bioessays* **35:** 780.

Loftus R.T., MacHugh D.E., Bradley D.G., Sharp P.M. and Cunningham P. (1994) *Proc. Natl. Acad. Sci. U S A* **91:** 2757.

Mirhoseinie S.Z., Vahidie S.M.F. and Gharehyazie B. (2005) *Iranian J. Biotech.* **3:** 41.

Rodriguez-Ramilo S.T., Fernandez J., Toro M.A., Hernandez D. and Villanueva B. (2015) *PLoS One* **10:** e0124157.

Stachowicz K., Sargolzaei M., Miglior F. and Schenkel F.S. (2011) *J Dairy Sci* **94:** 5160.

Valizadeh R., Nassiri M.R., Sadeghi B., Ghovati S.H. and Javadmanesh A. (2012) *Iranian J. of Anim. Sci. Res.* **3:** 444.

Waples R.S. (2006) *Conservation Genet.* **7:** 167.

Waples R.S. and Do C. (2008) *Mol. Ecol. Resour.* **8:** 753.

Weir B.S. and Cockerham C.C. (1984) *Evolution* **38:** 1358.

***ARTHROSPIRA PLATENSIS*: A NOVEL FEED SUPPLEMENT INFLUENCES GENE EXPRESSION IN THE HEART, KIDNEY AND LIVER OF PRIME LAMBS**

A.E.O. Malau-Aduli[1,2] and A. Kashani[1]

[1] Animal Sci & Genetics, TIA, School of Land & Food, University of Tasmania, Hobart, Australia
[2] College of Public Health, Medical & Veterinary Sci, James Cook University, QLD, Australia

SUMMARY

Transcriptional patterns in the expression of genes controlling lipid metabolism in supplemented sheep are currently unknown, thus quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) was used to assess the messenger RNA (mRNA) transcription levels and expressions of Aralkylamine N-acetyltransferase (*AANAT*), Adrenergic beta-3 receptor (*ADRB3*), B-cell translocation gene 2 (*BTG2*), and Fatty acid synthase (*FASN*) genes. The aim was to assess the effect of dietary supplementation with *Spirulina*, a blue-green cyanobacterial microalga, on the expression of genes controlling fatty acid metabolism in the heart, kidney and liver of prime lambs. In total, 432 heart, kidney and liver samples from 48 purebred and crossbred Merino prime lambs supplemented with low or high levels of *Spirulina* over a 9-week period were utilized for the study. Both the low and high levels of *Spirulina* supplementation regimes strongly up-regulated the transcription of all the selected genes. Sire breed and sex of lamb did not influence gene expression patterns; however, significant variations in response to *Spirulina* supplementation underpin the genetics-nutrition interactions that could be of practical importance for manipulating meat quality in the Australian dual-purpose prime lamb industry for a healthy polyunsaturated fatty acid profile.

INTRODUCTION

Spirulina (*Arthrospira platensis*) is a blue-green cyanobacterial microalga that contains 60-70% protein, high levels of carotenoids, essential vitamins, minerals and fatty acids (Ciferri, 1983, Holman and Malau-Aduli 2013; Holman *et al*. 2014). *Spirulina* has been trialled as a novel supplementary feed in many animal species and its recent inclusion in sheep diets has proven to be an effective nutritional strategy for increasing sheep meat production (Holman *et al*. 2012; Holman and Malau-Aduli, 2013). To our knowledge, apart from Kashani *et al*. (2015a) and Malau-Aduli and Kashani (2015), there is no available information on gene expression and molecular genetics-nutrition interactions between ovine organs and dietary supplementation with *Spirulina* in dual-purpose sheep. Therefore, the aim of this study was to investigate changes in the mRNA expression and transcriptional patterns of the following genes controlling lipid metabolism in the heart, kidney and liver of sheep under various *Spirulina* supplementation regimes: Aralkylamine N-acetyltransferase (*AANAT*), Adrenergic beta-3 receptor (*ADRB3*), B-cell translocation gene 2 (*BTG2*), and Fatty acid synthase (*FASN*).

MATERIALS AND METHODS

RNA was extracted from 432 heart, kidney and liver samples from 48 prime lambs sired by White Suffolk, Black Suffolk, Dorset and Merino rams randomly allocated to 3 treatment groups: the Control, grazing without *Spirulina* (0%), low (100mL/head/day in the ratio of 1g of *Spirulina* powder:10mL of water or 10% wt/vol), and high (200mL/head/day in the ratio of 2g of *Spirulina* powder:10mL of water or 20% wt/vol) *Spirulina* supplementation levels. The supplementary feeding trial continued for nine weeks after an initial three weeks of adjustment. The RNA samples were treated with PureLink™DNase (Life Technologies Pty Ltd. VIC, Australia), purified using RNeasy Mini Kit (Qiagen Ltd., VIC, Australia), DNase-treated and reverse transcribed to cDNA

using cDNA Synthesis Kit (Bioline Pty Ltd. NSW, Australia) and the primers in Table 1. Ubiquitin C (*UBC*) and Peptidyl-prolyl cis-trans isomerase A (*PPIA*) were used as house-keeping genes to normalise the gene expression data. Quantitative real time PCR *(qRT-PCR)* was carried out in triplicate 20 µL reactions containing 4 µL cDNA (50 ng), 10 µL 2×SensiFAST SYBR No-ROX Mix (Bioline Pty Ltd., NSW, Australia), 4.4 µL H_2O, and 0.8 µL forward and reverse primers (100 fmol). Assays were performed using the following cycling parameters: 95°C for 2 min (polymerase activation); 40 cycles of 95°C for 5 s (denaturation), 60°C for 10 s (annealing), and 72°C for 5 s (extension). Gene expression levels were recorded as cycle threshold (Ct) values, i.e. the number of PCR cycles at which the fluorescence signal is detected above the threshold value. Amplification efficiencies were determined for all candidate and reference genes using the formula E=10^(−1/slope), with the slope of the linear curve of cycle threshold (Ct) values plotted against the log dilution (Higuchi *et al*., 1993). The software package Rotor-Gene 3000 version 6.0.16 (Qiagen Pty Ltd., VIC, Australia) was used for efficiency correction of the raw Ct values, inter-plate calibration, normalisation to the reference gene, calculation of quantities relative to the highest Ct, and $\log_2$ transformation of the expression values for all genes. The qRT-PCR results were calibrated and normalized using the qBase relative quantification software (Pfaffl, 2001). A generalised linear model (GLM) in SAS (2009) was used in computing the fixed effects of *Spirulina* supplementation level, tissue, sire breed and sex, and their interactions on mRNA expression. Bonferroni's probability pairwise comparison test was used to separate mean differences, with the level of significance defined as $P<0.05$.

Table 1. Quantitative real-time PCR (qRT-PCR) oligonucleotide primers.

[a]Gene symbol	qPCR Primers		[b]T_a	Size (bp)
	Forward Primer	Reverse Primer		
AANAT	ACTGACCTTCACGGAGATGC	TTCACTCATTCTCCCCGTTC	60	211
ADRB3	TCAGTAGGAAGCGGGTCGGG	GGCTGGGGAAGGGCAGAGTT	60	291
BTG2	CTGGAGGAGAACTGGCTGTC	AAAACAATGCCCAAGGTCTG	60	194
FASN	GTGTGGTACAGCCCCTCAAG	ACGCACCTGAATGACCACTT	60	110
UBC	CGTCTTAGGGGTGGCTGTTA	AAATTGGGGTAAATGGCTAGA	60	90
PPIA	TCATTTGCACTGCCAAGACTG	TCATGCCCTCTTTCACTTTGC	60	72

[a]Aralkylamine N-acetyltransferase=*AANAT*, *β*3-adrenergic receptor=*ADRB3*, B-cell translocation gene 2= *BTG*2, Fatty acid synthase=*FASN*, Ubiquitin C=*UBC*, Peptidyl-prolyl cis-trans-isomeraseA=*PPIA*,
[b]T_a= Annealing Temp

RESULTS AND DISCUSSION

The relative mRNA expression levels of *AANAT*, *ADRB3*, *BTG2* and *FASN* genes in the heart are presented in Figure 1. Significant up-regulation of the *FASN* gene ($P<0.041$) in the heart, corresponding to a 21-fold change was observed in the low group. *AANAT* gene transcription was 4-fold and *BTG2* 1.4-fold up-regulated in the high treatment.

Kidney expression levels of *AANAT* and *ADRB3* genes increased by 16.64-fold and 54.53-fold respectively in the high group, while a down-regulation in the *BTG*2 transcript (0.15-fold) was observed in the low group (Figure not shown).

In the liver, high *ADRB3* and *BTG2* mRNA expression levels were detected in both the low or high groups. Significant up-regulation of the *ADRB*3 gene ($P<0.032$), corresponding to 60.59-fold in the low group, *BTG*2 gene ($P<0.024$), corresponding to 21.63-fold in the high group were noted (Figure not shown).

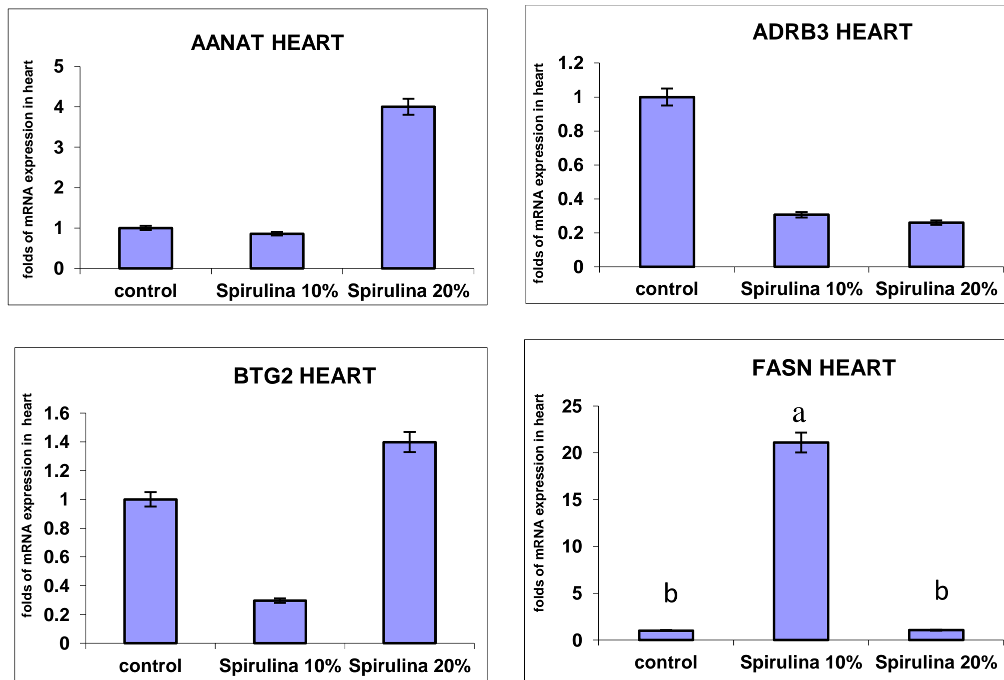


Figure 1. mRNA gene expressions in the heart of Australian prime lambs.

FASN is fundamental in enzyme regulation of the *de novo* synthesis step of lipogenesis and its main function is to catalyse the synthesis of palmitate from acetyl-CoA and malonyl-CoA, in the presence of NADPH, into long-chain FAs (Berndt *et al*., 2007; Byrne *et al*., 2005, Boizard *et al*., 1998). In the heart, the low level of *Spirulina* supplementation increased the transcription levels of the *FASN* gene leading to an increase in omega-6 (ω-6) and omega-9 (ω-9) PUFA and a decrease in saturated FA and a concomitant increase in PUFA in the heart of supplemented lambs.

In the kidney, significant expression of the *AANAT* encoded proteins accelerated the rate-limiting step in the synthesis of melatonin from serotonin (Coon *et al*., 1999, Reiter *et al*., 2014). Melatonin is a hormone that controls the function of the circadian clock, which regulates activity and sleep (Coon *et al*., 1999, Reiter *et al*., 2014). Spanish scientists discovered that melatonin consumption stimulated browning of white fat tissue in rats (Jiménez-Aranda *et al*., 2013). Brown fat burns, rather than stores, calories. Thus, melatonin has an anti-obesity effect and its metabolism protects against oxidative degradation of PUFA (Jiménez-Aranda *et al*., 2013, Reiter *et al*., 2014).In agreement, our fatty acid results (Kashani *et al*. 2015b) demonstrated that medium-level *Spirulina* supplementation significantly increased the ω-3 and ω-6 PUFA composition in all tissues and organs. The *ADRB3* gene plays a key role in regulating mammalian energy storage and expenditure (Malau-Aduli and Kashani 2015). It is also a principal mediator of the lipolytic and thermogenetic effects of high catecholamine (Forrest *et al*., 2007, Wu *et al*., 2011). The primary role of this receptor is in the regulation of resting metabolic rate and lipolysis (Forrest *et al*., 2003). Given that *ADRB3* encodes proteins regulating mammalian energy storage and expenditure by mediating effects from the sympathetic nervous system (Hu *et al*., 2010, Wu *et al*., 2012), our observations suggest an intricate genetics-nutrition interaction underpinning transcription at the molecular level that can be dietarily manipulated to achieve healthy FA composition outcomes.

BTG2 belongs to the anti-proliferative gene family and has been shown to be involved in cell growth, differentiation and survival (Mo *et al*., 2011), muscle fibre size, intramuscular fat deposition and weight loss (Kamaid and Giráldez, 2008, Sasaki *et al*., 2006). This seems to suggest that dietary *Spirulina* supplementation increases metabolic rate and lipolysis in the liver through up-regulation of *ADRB3* and simultaneously induces a decline in preadipocyte proliferation, an increase in energy expenditure, and a decline in energy uptake in adipocytes, ultimately enhancing ω-3 and ω-6 PUFA contents in the liver.

The results presented here demonstrated that mRNA expression levels of *AANAT*, *ADRB3*, *BTG2* and *FASN* in the heart, kidney, and liver are likely influenced by dietary *Spirulina* supplementation level. Our results show that genes involved in fatty acid metabolism in the kidney and liver are more sensitive to *Spirulina* supplementation than in the heart. These findings support the use of a low level of dietary *Spirulina* supplementation for optimal increase in healthy omega-3 and omega-6 fatty acid contents of organs among Australian crossbred sheep.

REFERENCES

Berndt J., Kovacs P., Ruschke K., Klöting N., Fasshauer M., *et al*. (2007) *Diabetologia* **50**:1472.

Boizard M., Le Liepvre X., Lemarchand P., Foufelle F., Ferré P. *et al*. (1998) *J. Biol. Chem.* **273:**29164.

Byrne K.A., Wang Y.H., Lehnert S.A., Harper G.S., McWilliam S.M., *et al*. (2005) Gene expression profiling of muscle tissue in Brahman steers during nutritional restriction. *J Anim Sci* **83:**1.

Ciferri O. (1983) Microbiol. Rev. **47:**551.

Coon S.L., Zarazaga L.A., Malpaux B., Ravault J.P., Bodin L., *et al*. (1999) *American J. Physiol. Endocrinol. Metab.* **277:**E792.

Forrest R.H., Hickford J.G.H. and Frampton C.M. (2007) *J. Anim. Sci.* **85:**2801.

Forrest R.H., Hickford J.G.H., Hogan A. and Frampton C. (2003) *Anim. Genet.* **34:**19.

Higuchi R., Fockler C., Dollinger G. and Watson R. (1993) *Biotech* **11:**1026.

Holman B.W.B., Kashani A. and Malau-Aduli A.E.O. (2014) *Small Rum. Res.* **120:**6.

Holman B.W.B. and Malau-Aduli A.E.O. (2013) *J. Anim. Physiol. Anim. Nutr.* **97:**615.

Holman B.W.B., Kashani A. and Malau-Aduli A.E.O. (2012) *American J. Expt. Agric.* **2:**160.

Hu J., Zhou H., Smyth A., Luo Y. and Hickford J.G. (2010) *Mol. Biol. Rep.* **37:**3389.

Jiménez-Aranda A., Fernández-Vázquez G., Campos D., Tassi M., Velasco-Perez L., *et al*. (2013) *J. Pineal Res.* **55:**416.

Kamaid A. and Giráldez F. (2008) *Dev. Dynamics* **237:**2158.

Kashani A., Holman B.W.B., Nichols P.D. and Malau-Aduli A.E.O. (2015a) *J. Anim. Sci. Tech.* **57:**8 DOI 10.1186/s40781-015-0047-3

Kashani A., Holman B.W.B., Nichols P.D. and Malau-Aduli A.E.O. (2015b) *Annals Anim. Sci.* **15:** DOI: 10.1515/aoas-2015-0037

Malau-Aduli A.E.O. and Kashani A. (2015) *Genes Genom* **37:** DOI 10.1007/s13258-015-0294-1

Mo XY, Lan J, Jiao QZ, Xiong YZ, Zuo B, et al. (2011) *Mol. Biol. Rep.* **38:**4389.

Pfaffl M.W. (2001) *Nucl. Acids Res.* **29:** e45.

Reiter R.J., Tan D.X. and Galano A. (2014) *Front. Physiol.* **5:**377.

SAS (2009) Statistical Analysis System. SAS Institute, version 9.2, Cary, NC, USA.

Sasaki Y., Nagai K., Nagata Y., Doronbekov K., Nishimura S., *et al*. (2006) *Anim. Genet.* **37:**40.

Wu J., Liu W., Liu J., Qiao L. and Yuan Y. (2011). *Anim.* **5:**88.

Wu J., Qiao L., Liu J., Yuan Y. and Liu W. (2012) *Mol. Biol. Rep.* **39:**8395.

ENVIRONMENTAL EFFECTS ON POST WEANING FLEECE TRAITS OF A MERINO SIRE EVALUATION FLOCK

S.I. Mortimer and A.E. Casey

[1] NSW Department of Primary Industries, NSW, Australia

SUMMARY

Significance of environmental and ewe bloodline effects, and their estimates, on post weaning (PW) fleece traits recorded in a central test sire evaluation flock were investigated. The findings that heavier fleeces were produced by single-born or single-reared or older animals, were generally similar to those reported for yearling expressions of clean and greasy fleece weights. Apart from sex significantly influencing each of the fleece quality traits, mean and coefficient of variation of fibre diameter, staple length and staple strength were largely unaffected by the environmental effects. Significant ewe bloodline and sire x ewe bloodline effects indicate that sire x genotype interactions are important for early age fleece traits. At present, these interactions are partially accounted for by fitting sire x flock interactions in the methodology applied in the MERINOSELECT genetic evaluations.

INTRODUCTION

A gradual decline in the average age at recording of performances in Merino fleece traits is evident in the MERINOSELECT database (D.J. Brown *pers. comm*). Therefore, understanding the importance of age effects on genetic parameters for fleece traits and the role of assessments in young Merinos in breeding programs designed to improve lifetime performance is of continuing interest. Performance recording of fleece traits by ram breeders is occurring more often now in an age range of 7 to 10 months (post weaning, PW), which offers ram breeders the opportunity to supply rams more in line with the needs of their commercial clients. As no published information is available on the genetics of Merino PW fleece traits, including the importance of environmental effects, this preliminary study sets out to estimate environmental effects on PW fleece traits recorded at a central test sire evaluation (CTSE) site, as well as the impact of ewe bloodline effects on these traits.

MATERIALS AND METHODS

Data for this study were available from the progeny of the Macquarie (Trangie) CTSE site, which is conducted according to the requirements of the Australian Merino Sire Evaluation Association (Casey *et al*. 2009). The site is hosted at the Agricultural Research Centre, Trangie. Briefly, the progeny were born in 2012 and 2013 following annual AI matings of 12 industry sires (with one sire in common across years) to base ewes sourced from commercial flocks representative of 2 Merino bloodlines (GRASS: dual purpose type, fleece weight emphasis; Karbullah: dual purpose type, fertility emphasis) and a commercial flock managed at the centre (progeny of ewes mated to Centre Plus rams: dual purpose type, wool, fertility and growth emphasis, 2013 mating only). Approximately 50 ewes were inseminated per sire. To achieve a spread of lambing in each year similar to that from natural mating, AI matings were conducted in 2 programs per year: 13 days apart in 2012 and 29 days apart in 2013. At their first mating, all ewes were maidens (aged 2 years old in both years, as well as 3 year old ewes in 2012) and were randomly allocated to each sire, balanced across the ewe bloodlines and AI programs within a year. Before lambing, the base ewes were allocated to a lambing paddock according to their AI sire and program (24 lambing paddocks per year). During daily lambing rounds, lambs were tagged and their date of birth and sex were recorded. Lambs were weaned at an average age of 92

days in 2012 and 115 days in 2013. Sire pedigree was confirmed and maternal pedigree assigned through DNA testing of blood samples from sires, dams and progeny conducted through the CRC for Sheep Industry Innovation. Birth type and rearing type of the progeny were derived from lambing paddock records, the DNA parentage testing results and dam's pregnancy scanning status.

Animals were first shorn at assessment of the PW fleece traits, which occurred at average ages of 8 months (2012 drop) and 8.6 months (2013 drop). Traits included greasy fleece weight (GFW, kg), clean wool yield (YLD, %), clean fleece weight (CFW, kg), mean fibre diameter (FD, micron), coefficient of variation of FD (FDCV, %), fibre curvature (CURV, degrees/mm), staple length (SL, mm) and staple strength (SS, N/ktex). Data analyses were performed using ASReml (Gilmour *et al*. 2014). Significance of fixed effects was first tested in models that fitted a random effect of sire. The fixed effects examined included ewe bloodline (3 levels), AI program group (4 levels), sex (ewe, wether), birth type (single, twin), rearing type (single, twin) and dam age (2, 3 and 4 year old matings). Age at observation was fitted as a linear covariate within each birth year. The importance of the interaction of sire with ewe bloodline was tested by fitting it as an additional random effect in a second model for each trait. The interaction effect was considered significant ($P<0.05$) if its inclusion with the sire effect resulted in a significant increase in the log-likelihood between models. Table 1 summarises the data on each trait.

Table 1. Descriptive statistics for post weaning fleece traits

Trait	Mean	SD	Range	Trait	Mean	SD	Range
GFW (kg)	3.4	0.73	1.4 - 6.0	FDCV (%)	18.1	1.89	13.0 - 25.7
YLD (%)	68.7	4.67	52.6 - 81.0	CURV (°/mm)	90.4	7.62	65.7 - 117.5
CFW (kg)	2.3	0.53	0.9 - 4.2	SL (mm)	66.6	9.15	40 - 96
FD (micron)	17.1	1.25	13.6 - 23.0	SS (N/ktex)	35.4	14.09	10.1 - 73.5

RESULTS AND DISCUSSION

Environmental effects. All PW fleece traits were significantly influenced by sex of the animal, with wethers having heavier (0.27 kg, greasy; 0.18 kg clean) and finer fleeces (0.26 micron) of slightly lower FDCV and CURV and with shorter staples of lower strength than ewes (Table 2). This contrasts with the study of Asadi Fozi *et al*. (2005), where no significant differences between ewes and wethers were found for GFW and CFW at 10 months of age of a multiple-bloodline fine wool flock; however, significant differences between the sexes were reported for YLD, FD, FDCV and SS. Earlier, Young *et al*. (1965) also observed no significant difference between rams and ewes in GFW at 5-6 months of age in a medium wool flock.

Age at observation significantly affected both fleece weights and SL, with older animals tending to produce more wool. The direction of the minor effect on SL varied with year of birth such that 2012-born animals (from a lambing period of 22 days) produced 0.09 mm shorter staples per day as age increased while 2013-born animals (from lambing period of 41 days) produced 0.33 mm longer staples per day (Table 2). In agreement with the current study, previous studies have reported significant effects of age at observation on GFW and CFW across a range of Merino wool types at 10 months of age (multiple-strain flock: Brash *et al*. 1997; multiple bloodline broad wool flock: Hill 2001; Asadi Fozi *et al*. 2005) and yearling age (industry flocks: Huisman *et al*. 2008) and SL (10 months of age: Hill 2001; yearling age: Huisman *et. al* 2008). However, these studies did show significant age effects on FD (Brash *et al*. 1997; Hill 2001; Asadi Fozi *et al*. 2005; Huisman *et al*. 2008) and SL (Hill 2001; Huisman *et al*. 2008).

Twin-born and twin-reared animals had significantly lighter PW GFW (0.19 and 0.26 kg respectively) and CFW (0.14 and 0.18 kg respectively; Table 2), which agrees with the significant effects on fleece weights previously identified in young Merinos due to birth type and rearing type

Table 2. Significance of fixed effects[1] and their estimates[2] for post weaning fleece traits and percentage of phenotypic variation accounted for by sire x ewe bloodline interaction

	GFW	YLD	CFW	FD	FDCV	CURV	SL	SS
Sex	***	**	**	***	**	***	***	*
	0.27	-0.78	0.18	-0.26	-0.31	-0.24	-4.09	-1.57
	(0.07)	(0.26)	(0.05)	(0.07)	(0.11)	(1.01)	(0.78)	(0.79)
Age	***	‡	***	n.s.	n.s.	n.s.	**	n.s.
2012	0.008	-0.27	-0.002				-0.09	
	(0.015)	(0.14)	(0.011)				(0.26)	
2013	0.034	-0.06	0.022				0.33	
	(0.006)	(0.05)	(0.004)				(0.10)	
Birth type	*	n.s.	*	n.s.	*	n.s.	n.s.	n.s.
	-0.19		-0.14		-0.74			
	(0.08)		(0.06)		(0.30)			
Rearing type	**	n.s.	**	n.s.	*	n.s.	n.s.	n.s.
	-0.26		-0.18		0.76			
	(0.08)		(0.06)		(0.29)			
Dam age	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.
3				-0.45				
				(0.16)				
4				-0.49				
				(0.22)				
Ewe bloodline	*	n.s.	*	***	*	***	***	n.s.
Karbullah	-0.09		-0.06	0.35	-0.32	-3.10	3.15	
	(0.03)		(0.02)	(0.08)	(0.12)	(0.51)	(0.73)	
Commercial	0.08		-0.02	-1.25	-0.19	0.43	-2.44	
	(0.09)		(0.07)	(0.21)	(0.34)	(1.42)	(1.99)	
Sire x ewe bloodline (%)	0 n.s.	0 n.s.	0 n.s.	4.0 *	2.8 *	1.6 n.s.	2.5 *	2.9 *

[1] ‡, P < 0.10; *, P < 0.05; **, P < 0.01; ***, P < 0.001; n.s., not significant (P ≥ 0.10).
[2] Estimates expressed as follows: ewe bloodline as the deviation from GRASS estimate of estimates for each of the Karbullah and Commercial bloodlines; sex as the deviation from ewe estimate of the wether estimate; birth type and rearing type as the deviation from estimates for single animals of estimates for twin animals; dam age as the deviation from estimates for 2 year old matings of estimates for later matings.

(Young *et al.* 1965; Brash *et al.* 1997; Huisman *et al.* 2008) and a combined birth-rearing type (Hill 2001; Asadi Fozi *et al.* 2005). While in the current study the estimates for rearing type effects on the fleece weights were as least as large as those for the birth type effects, Huisman *et al.* (2008) reported that the rearing type estimates for yearling fleece weights were approximately half those for birth type effects (-0.078 versus -0.193 for GFW and -0.074 versus -0.130 for CFW). It is possible that rearing type may have greater impacts on fleece weights recorded at the PW age than at yearling and later ages, but this needs to be confirmed by further study. FDCV was reduced significantly in fleeces of twin-born animals but was significantly higher in fleeces of twin-reared animals, as reported by Husiman *et al.* (2008). The remaining traits were not influenced by birth type or rearing type, results which are consistent with earlier findings for CURV (Huisman *et al.* 2008), SL and SS (Hill 2001; Huisman *et al.* 2008). However, birth type had influenced significantly FD in the studies of Brash *et al.* (1997) and Huisman *et al.* (2008), as well as a combined birth and rearing type effect in the studies of Hill (2001) and Asadi Fozi *et al.* (2005).

Except for FD, there were no significant differences between dam ages for the PW fleece traits, likely due to the ewes being no more than 4 years old at their second lambing opportunity. Based on published reports of significant estimates for fleece weights at 5-6 (Young *et al.* 1965) and 10-

12 (Hill 2001; Asadi Fozi *et al*. 2005; Huisman *et al*. 2008) months of age, it is expected that dam age effects could influence PW fleece weights. However, it is likely that these effects would be relatively small and less noticeable than effects of the other early environmental effects. This may also be the case for FD and FDCV, for which significant dam age effects, though small, have been reported (Hill 2001; Huisman *et al*. 2008).

Ewe bloodline effects. Ewe bloodline had a significant effect on all PW traits, except YLD and SS, with ranges of difference in performance across the ewe bloodlines used in this study of 0.17 kg for GFW, 0.06 kg for CFW, 1.6 micron for FD, 0.32% for FDCV and 5.59 mm for SL (Table 2).The sire x ewe bloodline interaction was significant only for FD, FDCV, SL and SS, accounting for 2.5 to 4.0% of the phenotypic variation in these traits and suggesting that non-additive genetic variance, or heterosis, may be influencing these traits. These results contrast with Mortimer *et al*. (1994) who found non-additive genetic effects to be significant for hogget GFW and CFW, but not hogget FD. From data drawn from industry studs and CTSE sites, Atkins *et al*. (1998) concluded that possible non-additive effects could be influencing yearling fleece weight, based on a significant sire x genetic group effect and the lesser importance of a sire x location effect following decomposition of a sire x management group interaction. Taken together, the evidence suggests that sire x genotype interactions are important for early age fleece traits, though further studies to estimate these effects are needed using suitable data. Currently, this interaction is in part being accounted for as a component of a sire x flock interaction that is fitted by the methodology used in MERINOSELECT genetic evaluations (Li *et al*. 2015).

Conclusions. Although to be verified by estimates from larger data sets, preliminary estimates of environmental effects on PW fleece traits are generally similar to those available for yearling fleece traits. The influence of sire x genotype interactions on early age fleece traits warrants further investigation using data sets structured without confounding of sire and other effects.

ACKNOWLEDGMENTS

Australian Wool Innovation Limited is acknowledged for funding this study. We are grateful for support offered to the Macquarie (Trangie) CTSE site by the ram breeders who entered rams in the 2012 and 2013 evaluations and the Australian Merino Sire Evaluation Association. Fleece data collection was ably conducted by T. Patterson, R. Rollinson, T. Bird-Gardiner, M. Brabrook, T. Atkinson, A. Jones, D. Trengrove, J. Meckiff, J. Lucas, B. Wilson and A. Russell.

REFERENCES

Asadi Fozi M., van der Werf J.H.J. and Swan A.A. (2005) *Aust. J. Agric. Res.* **56**: 789.

Atkins K.D., Gilmour A.R., Thompson R. and Coelli K.A. (1998) *Proc. 6th Wld. Congr. Genet. Appl. Livest. Prod.* **24**: 3.

Brash L.D., Taylor P.J. and Gilmour A.R. (1997) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **12**: 529.

Casey A.E., Martin S.J., Swain B.C. and Ramsay A.M. (2009) AMSEA accredited Central Test Sire Evaluation - Requirements for an accredited site. Edition 3.31219.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. and Thompson R. (2014) 'ASReml User Guide Relaese 4.0'. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Hill J.A. (2001) PhD Thesis, University of Adelaide.

Huisman A., Brown D.J., Ball A.J. and Graser H.-U. (2008) *Aust. J. Exp. Agric.* **48**: 1177.

Li L., Swan A.A., Brown D.J. and van der Werf J.H.J. (2015) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21**: This volume.

Mortimer S.I., Atkins K.D., Eissen J., van Heelsum A., Burns, A.M., *et al.* (1994) *Wool Tech. Sheep Breed.* **42**: 243.

Young S.S.Y., Brown G.H., Turner H.N. and Dolling C.H.S. (1965) *Aust. J. Agric. Res.* **16**: 997.

METHANE EMISSIONS ESTIMATED BASED ON MILK FATTY ACIDS OF JERSEY AND FLECKVIEH X JERSEY COWS IN A PASTURE-BASED SYSTEM

S. Abel[1], C.J.C. Muller[2] and B. Sasanti[1]
[1]Institute of Biomedical and Microbial Biotechnology, Cape Peninsula University of Technology, Bellville 7535, South Africa; [2]Western Cape Department of Agriculture, Research & Technology Development Services, Directorate: Animal Sciences, Private bag X1, Elsenburg 7607, South Africa

SUMMARY

Methane emissions by dairy cows have become a major environmental issue as it has been suggested as a major contributor to global warming. In this paper, the methane emissions of 40 Jersey and 62 Fleckvieh x Jersey (FxJ) cows were predicted using specific milk fatty acids (FAs). Cows were on kikuyu pasture supplemented during pasture shortages with a replacement mixture consisting of lucerne hay, oat hay and soybean meal. All cows received the same concentrate mixture twice a day after milking at 7 kg per cow per day. Milk samples (153 Jersey and 283 FxJ), were collected every 35 days over the lactation period, starting from 10 days after calving as per standard milk recording procedures. At least three and maximum seven milk samples per cow were collected over the lactation period. FAs were determined by gas chromatography. From this, methane emissions per unit of dry matter intake using two prediction equations were predicted. Predicted methane emissions varied between breeds although differences were small ($P>0.05$). Using Methane equation 2 and 3, predicted methane emissions for Jersey and FxJ cows were 26.2±1.07 and 25.8±0.94 and 24.6±0.99 and 24.4±1.08 g/kg feed DM, respectively. Methane emissions in both breeds showed curvilinear ($P<0.05$) trends over the lactation period. Over all cows, lactation stage and breeds, the coefficient of variation in predicted methane emissions for Jersey and FxJ cows was approximately 20%, possibly indicating genetic variation among cows. This variation could be used towards reducing methane production in dairy cows.

INTRODUCTION

Methane (CH_4) emissions by dairy cows are a major environmental issue as it has been suggested to be a major contributor to global warming. Dairy cows, being ruminants, contribute directly to greenhouse gas (GHG) emissions as CH_4, a major GHG, is produced through the fermentation processes in the rumen. Methane is released into the atmosphere by natural processes of eructation and breathing. Knapp *et al*. (2014) pointed out that to mitigate enteric CH_4 emissions per unit of energy-corrected milk, herd productivity, not individual animal productivity is important to environmental sustainability. This concept was demonstrated by Capper *et al*. (2009) in comparing US dairy production systems in 1944 and 2007 showing that while the carbon footprint of modern dairy cows increased from 13.5 to 27.8 CO_2 equivalents kg/cow, the CO_2 equivalent production decreased from 3.66 to 1.35 kg/kg milk. This resulted mainly from an improved genetic merit for milk production in dairy cows, better diet formulation, improved herd health and housing. The overall improvement in production resulted in 64% fewer dairy cows to produce the required milk output. Van de Haar & St Pierre (2006) and Chagunda *et al*. (2009) showed that more energy efficient dairy cows produce less methane and nitrogen per unit product. Selecting for dairy cows for more efficient feed use would bring together higher production and reduced resource requirements (Bell *et al*. 2012). However, by breeding for cows to produce less methane without forfeiting production would have a greater improvement on the industry's environmental foot-print. To enable this, a robust indicator trait that can be used together with production traits is required. Dijkstra *et al*. (2011) developed an indicator for methane production

of dairy cows based on milk FAs. However, correlations between methane production and individual milk FAs are not consistent among studies (Visker *et al*., 2014). Early work by Tyrrell *et al*. (1991) did not show breed differences between Holstein and Jersey cows for maintenance and production requirements per unit of metabolic body weight. Münger & Kreuzer (2006) also found no differences between Jerseys, Holsteins and Simmental in CH_4 per DM intake in a 3-three comparison with direct measurements of CH_4 and milk production, although Simmentals produced more CH_4/energy corrected milk because of a higher body size. The aim of this paper is to compare the predicted methane emissions of Jersey and FxJ cows in a pasture-based feeding system.

MATERIAL AND METHODS

Location and Animals. This paper was based on an on-going breed comparison study at the Elsenburg Research Farm of the Western Cape Department of Agriculture. Jersey and F bulls were used on two comparable groups of Jersey cows regarding breeding value for milk yield, to create a FxJ crossbred herd with a purebred Jersey herd as control. Cows were mostly on kikuyu pasture (*Pennisetum clandestinum*) being supplemented with a concentrate mixture twice a day after milking at 7 kg per cow per day regardless of milk yield and lactation stage. Fresh drinking water was freely available at all times. Milk samples, 153 from Jersey and 283 from FxJ, were collected over the lactation period starting from 10 days after calving every 35 days as per usual milk recording procedures. At least three and maximum seven milk samples per cow were collected over the lactation period. FAs were determined by gas chromatography. As the original aim of the study was the comparison of the production performance of the two breeds, cows were, at times, specifically during pasture shortages, supplemented with a pasture replacement mixture consisting of lucerne hay, oat hay and soybean meal. Results on the production performance of Jersey and FxJ cows have been reported by Goni *et al*. (2014).

Milk sampling and fatty acid analyses. Milk samples for FA analyses were collected and recorded every five weeks according to standard milk recording procedures. At each recording event, milk samples were collected from cows of both breeds from 10 days after calving (DIM) to about 175 DIM (milk tests 1 to 5) and thereafter from 240 DIM (milk tests 7 to 8). Milk samples were collected at the evening and next morning's milking session and combined. Milk samples for FA analyses were kept frozen at -20°C until laboratory analysis and FA composition of milk samples was obtained by gas chromatography at IBMB, CPUT, Bellville, Cape Town. Thirty six FA were detected and concentration levels determined. Breed differences for FAs were reported by Sasanti *et al*. (2012). For the present study methane emissions were predicted using the methane prediction formulae 2 and 3 as suggested by Visker *et al*. (2014):

CH_4 (g/kg feed DM) = 28.60 - 1.13 x C4:0 + 0.36 x C18:0 - 2.57 x C18:1*trans*10+11 - 9.29 x C18:1*cis* 11 for Methane 2.

CH_4 (g/kg feed DM) = 27.13 – 3.04 x C4:0 + 2.71 x C6:0 – 1.63 x C18:1*trans10+11* for Methane 3.

Statistical analyses. Predicted methane emissions (prediction models 2 and 3) were compared between breeds by analysis of variance using samples of all cows within breed as replicates using the GLM procedure (SAS Institute Inc.).

RESULTS AND DISCUSSION

The fat content in J milk was higher ($P<0.05$) than in FxJ milk, i.e. 4.09±0.85 and 3.87±0.74% (Table 1). Goni *et al*. (2014) found that although the fat percentage of the milk of FxJ cows was

lower (P<0.05) than J cows, the fat yield of FxJ cows was higher (P<0.05) than that of J cows because of a higher milk yield. The differences between breeds in methane production as predicted by methane prediction equations 2 and 3 were, however, small. Methane emissions found by Visker *et al*. (2014) using the same equations, were lower than the results in the present study, being 21.34±1.23 and 20.87±0.82 for methane prediction models 2 and 3, respectively. Similarly, as found by Visker *et al*. (2014), the present study also showed that for both Jersey and FxJ cows the methane production was lower for the Methane3 prediction model in comparison to Methane2. The coefficient of variation was high (*ca*. 20%) for both prediction models possibly indicating differences among cows within breeds. This variation could possibly be used to identify more efficient cows in terms of methane production.

Table 1. The mean milk fat content (%) and methane production based on fatty acids in the milk of Jersey (J) and Fleckvieh x Jersey (FxJ) cows in a pasture-based feeding system

Parameters	Fat %		Methane2		Methane3	
Breeds	J	FxJ	J	FxJ	J	FxJ
Mean (g/kg feed DM)	4.09[a]	3.87[b]	26.2	25.8	24.6	24.4
Standard deviation	0.85	0.74	1.07	0.94	0.99	1.08
Minimum	2.60	1.59	20.5	21.8	21.1	14.7
Maximum	7.80	6.49	28.5	28.4	26.3	28.7

[a,b]Values with different superscripts within column between breeds differ at P<0.05

The effect of lactation stage as indicated by milk test on methane production is shown in Figure 1. Methane production for both methane prediction equations showed curvilinear trends towards the end of the lactation similar as the fat percentage for both breeds.

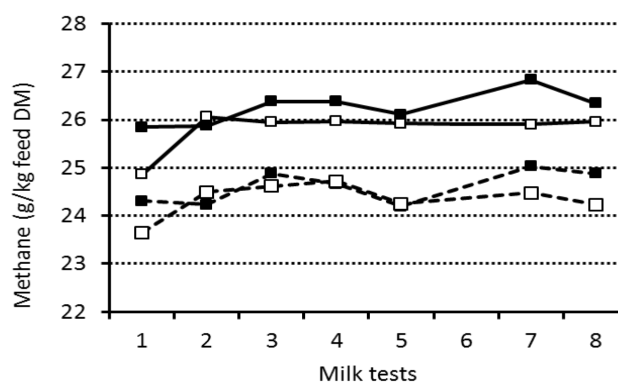


Figure 1. Methane production as affected by milk test for Jersey (■) and Fleckvieh x Jersey (□) cows using two prediction models (solid lines = Methane 2 and dash lines = Methane 3)

Although Garnsworthy *et al*. (2012) recorded methane emission rate during milking in automatic milking stations as CH_4/day, in contrast to the current study using predicted methane emissions as CH_4/kg DM, their study showed that between-cow variation in methane emission rate

was greater than within-cow variation. These differences between cows were ascribed to variations in body weight, milk yield, parity and week of lactation. Using daily methane emissions averaged on a weekly basis, Garnsworthy *et al*. (2012) showed an increase over the first 10 weeks of the lactation consistent with expected increases in feed intake because of milk yield increases. Lassen *et al*. (2013) demonstrated genetic variations with a heritability estimate of 0.21 using methane emission rates determined in automatic milking stations. High variation among cows increases the level of replication required to obtain reliable estimates of methane emissions and to assess the effect of mitigation strategies (Garnsworthy *et al*. 2012). Because milk is routinely collected for milk recording, FA profiles could easily be obtained using gas chromatography or infrared profiles (Visker *et al*. 2014). Regression equations developed by Visker *et al*. (2014) are probably not appropriate for the current study as diets differed. Chilliard *et al*. (2009) showed that CH_4 output was positively and strongly correlated to milk FA 6:0 to 16:0 based on the use of acetate produced in the rumen during fiber digestion. It was suggested that milk FA profile can be considered to be a potential indicator of *in vivo* CH_4 output, but studies using other dietary supplements reported contrasting results. Therefore, it appears that suitable equations should be developed for specific feeding programmes. For the present study it was, however, possible to show differences between cows within breeds.

CONCLUSION

This study showed that predicted methane emissions by using regression equations varied between cows within breed, while only small differences between Jersey and FxJ cows were observed. Within breed differences indicate possible genetic variation among cows which may be exploited towards reducing methane production in dairy cows.

ACKNOWLEDGEMENTS

This research was partly funded by the Western Cape Agricultural Research Trust. The efforts of the management team at the Elsenburg dairy are greatly appreciated.

REFERENCES

Bell M.J., Eckard R.J. and Pryce J.E. 2012. *http://dx.doi.org/10.5772/50395*.

Capper J.L., Cady R.A. and Bauman D.E. (2009). *J. Anim. Sci*. **87**: 2160.

Chagunda M.C.G., Römer D.A.M. and Roberts D.J. (2009) *Livest. Sci*. **122**: 323.

Chilliard Y., Martin C., Rouel J. and Doreau M. (2009) *J. Dairy Sci*. **92**: 5199.

Dijkstra J. Van Zijderveld S.M., Apajalahti J.A., Bannink A., Gerrits W.J.J., Newbold J.R., Perdok H.B. and Berends H. (2011) *Anim. Feed Sci. Technol*. **166**: 590.

Garnsworthy P.C., Craigon J., Hernandez-Medrano J.H. and Saunders M. (2012) *J. Dairy Sci*. **95**: 3181.

Goni S., Muller C.J.C., Dube B. and Dzama K. (2014) Trop. Anim. Health Prod. **46**: *DOI 10.1007/s11250-014-0698-y*.

Knapp J.R., Laur G.L., Vadas P.A., Weiss W.P. and Tricarico J.M. (2014). *J. Dairy Sci*. **97**: 1.

Lassen J. and Løvendahl P. (2013) *Adv. Anim. Biosci*. **4**: 280.

Münger A. and Kreuzer M. (2006). *Int. Congr. Ser*. **1293**: 119.

Sasanti B., Abel S., Muller C.J.C., Gelderblom W.C.A., Swanevelder S. and Schmulian A. (2012) *45th SASAS Congress*. p.110.

Tyrrell H.F., Reynolds C.K. and Blaxter H.D. (1991). *EAAP Publ.58*.

Van de Haar M.J. and St Pierre N. (2006) *J. Dairy Sci*. **89**: 1280.

Visker M.H.P.W., van Engelen S., Dijkstra J. van Arendonk J.A.M. and Bovenhuis H. (2014) *Proc. 10th Wo. Con. Gen. App. Livest. Prod*. **10**: 305.

COMPARISON OF SURVIVAL PHENOTYPES IN SPECIFIC DAIRY PRODUCTION SYSTEMS

B. Visser[1], K. Stachowicz[1], J.R. Bryant[2] and P.R. Amer[1]
[1]AbacusBio Limited, PO Box 5585, Dunedin
[2]DairyNZ Limited, Private Bag 3221, Hamilton

SUMMARY

Over thirty years' worth of dairy cattle data was analysed to help understand the phenotypic relationships between early (surviving from 1st parity to 2nd parity) and late (surviving from 1st parity to 4th parity) survival and other phenotypes in high and low milk production output systems on New Zealand dairy farms.

Survival and fertility traits were computed from parturition, herd test and culling records, and mating and parturition records respectively.

The results of this study indicate some phenotypic differences in relationships for Traits Other than Production (TOP), body condition score, Somatic Cell Score (SCS) with survival between farming systems and over time. Non-linear relationships between TOP traits and survival were observed. There was a tendency for SCS to influence survival more at a younger age than at an older age across all production levels and for Farmer Opinion to influence survival more at a younger age than at an older age within high output systems.

INTRODUCTION

New Zealand dairy production systems have become more diverse over the years with a higher proportion of farms becoming more intensive. This work is a preliminary analysis of low and high output production systems to help understand if there are different phenotypic and genotypic drivers of cow survival (longevity) to optimise trait selection.

This paper describes the phenotypic relationships between survival and production and non-production traits in low and high milk solids production systems. Phenotypic correlations can reflect trade-offs between traits and potentially reveal underlying genetic relationships.

The specific objective was to compute phenotypic correlations for early and late survival with Traits Other than Production (TOP) scores, 270 day combined milk solids, somatic cell count (SCC), liveweight, body condition score (BCS) and fertility. Comparisons were made across herd-level groupings based on two different production systems: high output versus low output, as well as two time periods (2002 to 2014 for all records and 2010 to 2014 for recent records). All phenotypes were based on first lactation records, except for survival which was based on first to fourth lactation records.

METHODS

Data was extracted from the NZ Dairy Core Database and Dairy Industry Good Animal Database (DIGAD). Data preparation and analysis was undertaken using SAS software (SAS Institute 2011) and R 3.1.2 (R Core Team 2014; Dowle *et al.* 2014). Pairwise Pearson correlations were computed between phenotypes from herds belonging to different herd-level subsets based on milksolids (fat plus protein) production in second lactation and record date.

Data preparation. Only herds with good quality data were selected. Animals had to: have birth date recorded (not estimated), be born in New Zealand, have been born after the year 2000 and have both parents recorded. Animals that moved herd and did not reside in a single herd from birth to their second lactation were discarded. Only herds with more than 50% of animals with TOP

scores were selected to obtain a representative sample of herds that had a high proportion of cows scored for TOP traits (Table 1).
Survival and fertility traits were computed from parturition, herd test and culling records, and mating and parturition records respectively.

Traits Other than Production scores included scores for milking adaptability and speed, temperament, farmer opinion, dairy capacity, rump angle and width, legs, udder support, front and rear udder, front and rear teat placement, udder overall and dairy conformation.

Table 1: Average first lactation milksolids (fat + protein) production, number of herds, animals in groups for which phenotypic correlations were assessed.

Production category	Milksolids production (kg/cow/day)	Herds	Animals All (2002-2014)	Animals Recent (2010-2014)
High output	1.38	34	13,058	4,886
Low output	0.96	42	14.199	4,024

Survival. Early and late survival traits were computed as per the existing multi-trait genetic evaluation (Harris *et al.* 2007), which defines survival as the ability of the cow to be retained in the herd consecutively from her first to second and first to fourth lactation (SV12 and SV14 respectively). While survival from first to third lactation is also included in the existing genetic evaluation, results for this trait have been omitted for conciseness. Parturition and herd test dates were used to determine whether a cow was lactating in a season, starting with the first lactation from one and a half years old to exclude heifers that calved early.

The lactation of a cow in a season was scored as a logical variable: 1 when they had a lactation in a season and 0 otherwise. When a cow had no lactation record(s) in between two seasons in which she was lactating, that lactation was defined as missing. Survival phenotypes were then derived by checking all logical lactation values over the interval from first to second and first to fourth lactations for SV12 and SV14 respectively. Survival was set to 1 if all the lactation values were 1. Survival was 0 otherwise.

Fertility. Mating and calving records were used to derive calving season day (CSD). Calving season day was calculated as the difference in days between planned start of calving (Stachowicz *et al.* 2014) and the cow's calving date.

Only spring calvings were used to compute CSD, i.e. calvings between June and November. Mating had to occur within 200 days from the previous calving. Calving intervals (difference in days between consecutive calvings) from 300 to 550 days were accepted. Mating dates where the individual bull could be identified were used and "run with bull" matings were excluded. For both mating and calving dates there had to be matching herd identification records.

Survival records were merged with fertility records to determine if a fertility-related culling event occurred for a cow during her first lactation. Cows that were mated but did not calve as a result of this mating were penalized by having 21 days added to the longest calving season day record in their contemporary group (Donoghue *et al.* 2004).

Groups. For the purpose of herd-level grouping, every animal was assigned to a single herd: the herd in which it calved for the first time. The majority of animals were not transferred during their lifetime, therefore the impact of assigning the smaller part of an animal's data to the wrong herd was assumed to be minor.

To allocate herds to a particular production system (high versus low output), phenotypes for daily combined milk solids production from second lactation herd test records recorded between

50 and 200 days in milk (lactation curve shows perfect linear fit (R^2=0.99) in this interval) were used. PROC GLM in SAS was utilized to fit a linear model including fixed effects of days in milk, herd, season, age of the cow at calving, month of calving and the cow's breed composition. High and low output herds were defined as those herds belonging to the top and bottom performers of the obtained herd lactation solution. Low and high solution thresholds were adjusted in such a way that there were approximately 20,000 animals in both groups. Records were also grouped by data age: recent (records from 2010 to 2014) and all (records from 2002 to 2014).

Analysis. Pearson correlation coefficients were computed among raw phenotypes, focussing on the correlations of the two survival traits (1st to 2nd lactation, and 1st to 4th lactation) with TOP and production traits. Linear and quadratic coefficients for the regressions of survival traits on the TOP traits were also estimated.

RESULTS AND DISCUSSION

There was a 0.42 kg difference in milksolids per cow per day between low and high output herds (Table 1). A summary of results is given in Table 2. In general, udder conformation traits had stronger phenotypic correlations in high rather than in low output herds with some evidence that the relationship between udder traits and early survival is becoming stronger in the more recent data. In high producing herds, overall farmer opinion, had a stronger correlation with early rather than with late survival. This corresponds with the earlier finding that farmer opinion has an important influence on survival (Berry *et al.* 2005). The phenotypic relationship between 270-day milk solids production and cow survival was constant across herd types and time periods. Somatic cell score had a strong negative relationship with early survival, and was stronger in high output herds than in low output herds. Milking adaptability showed a similar pattern, perhaps reflecting the correlation of approximately 0.6 observed between farmer opinion and milking adaptability in these data. There was a stronger phenotypic correlation between both udder support and udder overall and survival in high than in low output herds.

First lactation BCS had a low negative correlation with SV12 in high output herds. This might reflect that cows which maintain condition produce less milk in high output herds and this impacts on their likelihood of survival. Whereas in low output herds, it is more beneficial to maintain condition in early lactation as this confers benefits for fertility and late season milk production. Dairy capacity exhibited a low genetic relationship with survival in low production herds, which supports the above explanation as capacity and BCS are highly correlated.

Regression analysis results are not shown here in detail, but TOP scores had an impact on most survival in a non-linear manner, such that quadratic regression coefficients were significantly different from zero. Of 252 TOP/survival relationships studied, 142 were significant, and 122 out of 142 had significant non-linear phenotypic relationships, usually with an intermediate optimum.

Future research will estimate genetic correlations between survival traits, and all other traits by production level. The results of this study indicate that TOP and other trait records in daughters might more accurately inform survival genetic evaluations of dairy cattle if account is taken for the differences in relationships for TOP traits and other traits with survival across farming systems; by considering non-linear relationships and relationships specific to age. This later point is especially relevant because survival in early lactations is more economically important than survival across lactations at later ages. This is because the increased need to replacements is much higher in a herd where more cows are lost at young ages than at later ages. Finally, there is some evidence that the relationships between TOP traits and survival are changing over time, which has implication for the use of historic data for estimating genetic correlations.

Table 2 Phenotypic correlations between survival traits and TOP scores, 270 day combined milk solids, SCS, liveweight and BCS, and CSD for herd-level groupings based on milk solids production and data age.

Trait		Group							
	Production system	High output				Low output			
		All		Recent		All		Recent	
		SV12	SV14	SV12	SV14	SV12	SV14	SV12	SV14
TOP	Milking adaptability	0.05	0.03	0.05	0.07	0.03	0.01	0.07	0.07
	Temperament	0.05	0.03	0.02	0.06	0.04	0.01	0.07	0.08
	Milking speed	0.03	0.04	0.01	0.04	0.02	0.04	0.10	0.13
	Farmer opinion	0.13	0.07	0.13	0.07	0.07	0.06	0.09	0.11
	Dairy capacity	0.01	0.01	0.04	-0.07	0.03	0.05	-0.02	-0.05
	Rump angle	-0.02	0.02	-0.01	-0.05	0.00	0.01	-0.01	-0.07
	Rump width	0.02	0.03	0.02	0.00	0.03	0.05	0.00	0.04
	Legs	0.00	-0.02	0.01	-0.03	0.01	0.02	-0.01	0.02
	Udder support	0.08	0.08	0.12	-0.01	-0.01	0.03	-0.10	-0.08
	Fore udder	0.04	0.07	0.07	-0.01	0.02	0.04	-0.02	-0.04
	Rear udder	0.06	0.05	0.10	-0.05	-0.03	0.00	-0.12	-0.07
	Front teat placement	0.02	-0.05	0.04	-0.07	0.00	-0.03	0.02	-0.05
	Back teat placement	0.04	-0.04	0.04	-0.03	-0.01	-0.04	-0.02	-0.06
	Udder overall	0.08	0.08	0.11	-0.02	0.01	0.04	-0.05	-0.05
	Dairy conformation	0.08	0.07	0.10	-0.05	0.06	0.09	-0.02	-0.03
Liveweight		-0.02	0.02	-0.03	-0.02	0.01	0.01	-0.01	0.01
BCS		-0.07	-0.01	-0.10	-0.01	0.01	0.00	0.00	-0.05
270-day milk solids		0.11	0.11	0.10	0.09	0.12	0.13	0.08	0.08
SCS		-0.22	-0.07	-0.27	-0.06	-0.11	-0.05	-0.13	-0.03
CSD			0.00		0.01		-0.03		0.03

ACKNOWLEDGEMENTS

This work was funded by NZ Dairy farmers through the milk solids levy.

REFERENCES

Donoghue, K.A., Rekaya, R. and Bertrand, J.K. (2004) *J. Anim. Sci* **82**: 357.

Berry D.P. Harris B.L. Winkelman A.M. and Montgomerie W.A. (2005) *J Dairy Sci* **88**: 2962.

Harris, B.L. and Montgomerie, W.A., (2007) *Proc. New Zeal. Soc. Anim. Prod.* **67**: 377.

SAS Institute (2011) 'The SAS system for Windows. Release 9.2'.

Dowle M., Short T. and Lianoglou S. (2014) 'Data.table: Extension of data.frame'.

R Core Team (2014) 'R: A Language and Environment for Statistical Computing'.

Stachowicz K., Jenkins G. M., Amer P. R., Meier S. (2014) *Interbull Bulletin* **48**: 1.

MODELLING OF LONGITUDINAL LIVEWEIGHT DATA USING REGRESSION WITH LEGENDRE AND EIGENVECTOR FUNCTIONS

M. D. Price, M. E. Olayemi and J. B. Bryant

DairyNZ Limited, Private Bag 3221, Hamilton 3240, New Zealand

SUMMARY

Liveweight data from walk-over-weigh technology on NZ dairy farms provides a potential source of daily or average liveweight data for genetic selection. In this study, mixed models utilising both fixed and random regression with varying orders of Legendre polynomials or eigenvectors were assessed to model the longitudinal properties of liveweight records of New Zealand dairy cows. Higher order models fitted the data best despite the subsequently large increase in number of parameters. The choice of either Legendre polynomials or eigenvectors as the set of basis functions made no significant difference. Limits to model order will need to be applied on the basis of assumptions concerning data covariance structure and desired number of liveweight-derived traits for selection. Additionally, methods developed here for estimation of genetic covariance of regression coefficients and for inversion of the mixed model equation matrix can be extended to models for which pedigree relationships are also taken into account.

INTRODUCTION

Liveweight (LW) is a trait which changes over time, and may thus be considered as "longitudinal" or "infinite-dimensional" (Kirkpatrick *et al.*, 1994). Some of the approaches that have been applied to model these repeated data types include repeatability, multivariate and random regression (RR) models (Mrode and Thompson, 2014).

The use of RR models has become a preferred method to analyse longitudinal data for animal genetic evaluation. These models use a fixed regression to describe the average shape of a lactation or growth curve, and a random regression for each animal to account for deviations from the fixed regression (Schaeffer and Deckers, 1994). RR models have been applied to model test-day records of milk traits in dairy cattle (Jamrozik and Schaeffer, 1997; Olori *et al.*, 1999), as well as growth and mature weight in beef cattle (Meyer, 1999, 2004; Speidel *et al.*, 2010) and other livestock. In these models, Legendre polynomials are typically used as the set of basis functions, however they can suffer from a Runge effect (Runge, 1901), where a higher order polynomial describing the general curve has high oscillations in the boundary areas. Eigenvectors avoid this problem, and just a few eigenvectors may adequately account for the covariance structure of the data, so they may provide a better set of basis functions in a RR model.

The current animal evaluation model for LW of NZ dairy cattle is based on a combination of both visual score and static weights. Walk-over-weigh (WOW) records per animal in commercial dairy sheds provide the opportunity to incorporate this new data into the current LW models. Utilisation of these data may also allow for better characterisation of the seasonal LW curve. The objective of this study was to compare RR models using either Legendre or eigenvector basis functions of various orders with longitudinal LW data, albeit without considering pedigree.

MATERIALS AND METHODS

Data. A total of N = 58,532 WOW records collected on n_a = 2,899 2-year-old New Zealand dairy cows born in the 2010/2011 season were extracted from the Dairy Industry Good Animal Database (DIGAD). The data consisted of individual animal weekly average LW collected over n_t = 40 weeks from lactation start (defined as weeks-in-milk, WIM), from n_{cg} = 6 herds. All data manipulation, modelling and analysis were performed with R statistical software.

Models. Mixed models were used in the analysis, whose fixed component included a regression on either Legendre or eigenvector basis functions which would describe the general liveweight curve, and whose random component described individual deviations from the general curve. Following the notation of Mrode & Thompson (2014), models used were of the form:

$$y_{tij} = htd_i + \sum_{k=0}^{n_f} \psi_{tk}\beta_k + \sum_{k=0}^{n_r} \phi_{tk}a_{jk} + e_{tij} \quad (1)$$

Where y_{tij} is the test day record for cow j on day t within contemporary group (herd test day) i, ψ_{tk} is the value of a kth basis function (Legendre or eigenvector) evaluated at time t, $\beta_k \in \boldsymbol{\beta}$, the vector of fixed regression coefficients, ϕ_{tk} is the value of a kth Legendre polynomial evaluated at time t, and $a_{jk} \in \mathbf{a}_j$, the vector of random regression coefficients (animal effects) for animal j (where $\mathbf{a}_j \in \mathbf{a}$, the full vector of random regression components). The set of basis functions for fixed effects are of order n_f, and for random effects are of order n_r. The matrix notation (2) and mixed model equation (MME) notation (3) for this model are as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \quad (2)$$

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{I}\sigma_e^2 \otimes \mathbf{K}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix} \quad (3)$$

Here it is assumed that var($\mathbf{a}$) = $\mathbf{I}\otimes\mathbf{K}$ and var($\mathbf{e}$) = $\mathbf{I}\sigma_e^2$ (necessary priors), where $\otimes$ is the Kronecker product and $\mathbf{K}$ is an n_r-dimensional covariance structure of the random regression coefficients for animal effects. $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices corresponding to the effect solutions (superscript T indicates matrix transpose). Here $\mathbf{Xb} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\boldsymbol{\beta}$, the sum of contemporary group fixed effects ($htd_i \in \mathbf{b}_1$) and fixed regression components respectively. It should be noted that our approach does not yet take into account the pedigree relationships, which would otherwise partition the animal effect into additive animal genetic and permanent environmental effects. If this were the case, the MME would be in a 3×3 partition, with n_r-dimensional estimated covariance structures of $\mathbf{G}$ and $\mathbf{P}$, in Kronecker product with $\mathbf{A}^{-1}\sigma_e^2$ and $\mathbf{I}\sigma_e^2$, respectively. As it stands, we instead used an estimate of the combined covariance of $\mathbf{G}$ and $\mathbf{P}$; $\mathbf{K}$, which was the genetic covariance of animal effects $\text{cov}_A(\mathbf{a})$ calculated from the phenotypic covariance of the data $\text{cov}_P(\mathbf{y})$, with $\mathbf{y}$ structured as an animal × test-week dataset. Liveweight heritability was assumed to be $h^2 = 0.35$; residual variance was assumed to be σ_e^2 = 400kg.

$$\mathbf{K} = \text{cov}_A(\mathbf{a}) = \frac{h^2}{2}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\text{cov}_P(\mathbf{y})((\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)^T \quad (4)$$

Basis function sets. For Legendre polynomials, the setup of an order k incidence matrix (for a basis set of k functions for either the fixed or random model component) was calculated as $\mathbf{X}$ (or $\mathbf{Z}$) = $\mathbf{ML}$, where $\mathbf{L}$ was the $k\times k$ matrix of Legendre polynomial coefficients, and $\mathbf{M}$ was the $N\times k$ matrix of each observation's week-in-milk t transformed to the [-1, 1] interval (5) and evaluated for the k different degrees of monomials.

$$x_m = \frac{2(t_m - t_{min})}{(t_{max} - t_{min})} - 1, \qquad m \in \{1 \cdots N\} \quad (5)$$

For eigenvectors, the setup of an order k incidence matrix was calculated as $\mathbf{X} = \mathbf{TE}_k$, where $\mathbf{T}$ was the $N\times n_t$ matrix for each observation's WIM, and $\mathbf{E}_k$ was the $n_t\times k$ subset of the k eigenvectors of the top k eigenvalues of the eigendecomposition of the phenotypic covariance matrix, $\text{cov}_P(\mathbf{y}) = \mathbf{E\Lambda E}^T$ ($\mathbf{E}$ is the matrix of all eigenvectors; $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues).

Residual analysis. Akaike information criteria (AIC) (Akaike, 1974) were calculated for each model to provide a relative measure of model quality. For nested models, Likelihood ratio tests (LRT) (proven by the Neyman–Pearson (1933) lemma to be optimal for model selection) were also used to determine if any reduction in residuals between models was significant or not.

Assuming normal distribution of residuals, the log-likelihood of a model was determined as a function of residual variance σ^2 and number of observations N:

$$\ell n(\mathcal{L}) = -\frac{N}{2}\left(1 + \ell n(2\pi\sigma^2)\right) \qquad (6)$$

From this, AIC = $2n$ - $2\ell(\mathcal{L})$ was calculated (where $n = n_{cg} + n_f + n_r.n_a$ was the total number of parameters), and for two nested models a likelihood ratio LR = $2\ell n(\mathcal{L}_{M2})$ - $2\ell n(\mathcal{L}_{M1})$ was calculated (for "null" model M1 nested within model M2). A chi-squared test using test statistic LR with degrees of freedom $df = n_{M2} - n_{M1}$ would produce a value $p = 1 - \chi^2$ which, for $p < 0.05$, would indicate that model M2 was significantly better than the simpler M1 model.

MME matrix inversion. The block-structure of the MME matrix (dimension $n_{cg} + n_f + n_r.n_a$) allowed for an alternative inversion of much smaller n_r-dimensional matrices. If the data is ordered by animal, then the $\mathbf{Z}^T\mathbf{Z}$ component of the MME (3) will be block-diagonal, comprised of n_a sub-matrices of dimension n_r each. The same is true for $\mathbf{I}\sigma_e^2 \otimes \mathbf{K}^{-1}$, and so $\mathbf{D} = \mathbf{Z}^T\mathbf{Z} + \mathbf{I}\sigma_e^2 \otimes \mathbf{K}^{-1}$ will be block-diagonal also. Given the Banachiewicz (1937) identity for of a partitioned matrix inverse,

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{S}_D^{-1} & -\mathbf{S}_D^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{S}_D^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{S}_D^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}, \quad \text{where } \mathbf{S}_D = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \qquad (7)$$

and the property of the inverse of a block-diagonal matrix being another block-diagonal matrix of individual block inverses, then it follows that the inverse of the MME matrix may simply be determined by way of calculating the inverse of each n_r-dimensional sub-matrix of **D** (and the inverse of the small (n_{cg}+n_f)-dimensional Schur matrix $\mathbf{S}_D$).

RESULTS AND DISCUSSION

Varying model orders. For RR models with fixed Legendre component of $n_f \in \{3, \ldots, 7\}$ and random component of $n_r = 3$, the general polynomial representing the data (Figure 1) shows how the oscillations of higher order polynomials become apparent for $n_f \gtrsim 5$. Despite this, LRT's between models of consecutive orders of fixed Legendre component $n_f \in \{3, \ldots, 20\}$ with conserved random order $n_r \in \{1, \ldots, 3\}$ showed significantly better fit for orders $n_f \leq 9$ ($p < 0.005$). However, residual variance was not longitudinally homogenous (even for varying random orders), indicating that higher order fixed effects may be required. Unlike the random regression component, increasing the fixed order makes almost no computational difference. However, under the assumption of a reasonably smooth longitudinal relationship between LW values, the fixed order should be limited.

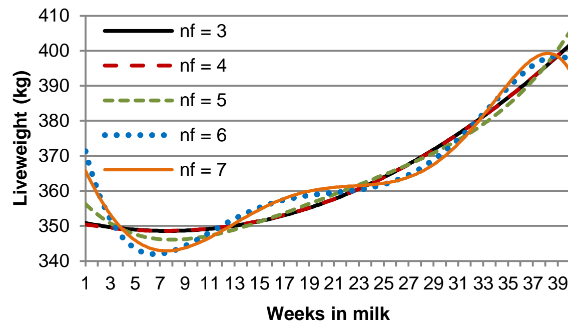


Figure 1. Fixed regression curves for models with Legendre orders $n_f \in \{3, \ldots, 7\}$ and $n_r = 3$.

For RR models with fixed Legendre component of $n_f = 9$ and random component of $n_r \in \{1, \ldots, 5\}$, residual variance decreases (and log-likelihood increases) with increasing n_r (Table 1), and this variance also scales quite uniformly across season for different n_r. LRT's indicated that models were very significantly better with increase in random component order for $n_r \leq 3$, for both Legendre and eigenvectors models. The AIC was minimised for $n_r = 3$, indicating that quadratic random animal effects are best. Increased parameters per animal should contribute in a biologically meaningful way, as it is upon those parameters that animals may be selected by. Therefore, while incorporating a random regression component into the model is advisable, the order of this component should not be too large; letting $n_r = 3$ should be sufficient.

Table 1. AIC and LRT's of models of Legendre or eigenvector fixed component $n_f = 9$, and $n_r \in \{1, \ldots, 5\}$

	Legendre, $n_f = 9$					Eigenvector, $n_f = 9$				
Order n_r	1	2	3	4	5	1	2	3	4	5
$\ell n(\mathcal{L})$ ($\times 10^3$)	-289	-280	-276	-273	-271	-289	-280	-276	-274	-272
AIC ($+5.7\times 10^5$)	13,904	2,533	449	1,051	2,895	14,661	2,930	568	1,217	3,256
LR (n_r vs. n_{r-1})		17,169	7,882	5,196	3,954		17,529	8,160	5,148	3,760
$p = 1\text{-}\chi^2$		0	0	1	1		0	0	1	1

Legendre vs. eigenvector. Models with a fixed regression component of either Legendre polynomials or eigenvectors of order $n_f = 9$ were compared via AIC values for their relative merit, for random component of Legendre orders $n_r \in \{1, \ldots, 5\}$. For any particular random order n_r, Legendre models had slightly better AIC values than those of the eigenvector models (Table 1). In the absence of a measure of significance for AIC comparison, the default choice of regression function ought to remain as the Legendre polynomials.

Pedigree information. Future inclusion of pedigree information ought to improve model fit even more due to increased utilisation of data via pedigree linkages. While the relationship matrix **A** would not be subject to inversion by the block-matrix inversion technique, the use of $\mathbf{A}^{-1}$ would allow for a similar technique for solving the model. Therefore the use of a full (including pedigree) RR model will have essentially no more computational complexity for random orders $n_r > 1$.

CONCLUSION

In a RR model for WOW data, increasing orders of fixed and random regression components significantly improve the model in general, though these must be tempered by the practical realities of assumed longitudinal relationships and necessary number of parameters per animal. The choice of type of regression function (Legendre polynomials or eigenvectors) is insignificant.

The future inclusion of pedigree relationships ought to ensure a much better depth of data per animal and subsequent model improvement. The use of block-matrix inversion will still ensure that computational complexity is significantly reduced.

REFERENCES

Akaike H. (1974) *IEEE Transactions on Automatic Control* **19**(6): 716.

Banachiewicz T. (1937) *Acta Astronomica, Series C*, **3**: 41.

Jamrozik J. and Schaeffer L. R. (1997) *J. Dairy Sci*. **80**: 762.

Kirkpatrick M., Hill W.G. and Thompson R. (1994) *Genet. Res.* **64**: 57.

Meyer K. (1999) *J. Anim. Bree. Genet.* **116**: 181.

Meyer K. (2004) *Livest. Prod. Sci*. **86**: 69.

Mrode R.A. and Thompson R. (2014) '*Linear Models for the Prediction of Animal Breeding Values', 3rd Ed. Cabi Publ., Wallingford, UK*. pp130–155.

Neyman J. and Pearson E.S. (1933) *Phil. Trans. Royal Soc. A: Math. Phys. Eng. Sci.* **231**(694–706):289.

Olori V.E., Hill W.G., McGuirk B.J. and Brotherstone S. (1999) *Livest. Prod. Sci*. **61**: 53.

R Core Team (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. URL http://www.R-project.org/

Runge C. (1901) *Zeitschrift für Mathematik und Physik* **46**: 224.

Schaeffer L.R. and Dekkers J.C.M. (1994) *Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Canada* **18**: 443.

Speidel S.E., Enns R.M. and Crews D.H. Jr. (2010) *Genet. Mol. Res*. **9**: 19.

FIBRE DIAMETER MEASURED IN THE POST-WEANING AGE WINDOW IS GENETICALLY THE SAME TRAIT AS YEARLING FIBRE DIAMETER

Andrew Swan[1,2]
[1]Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351
[2]Australian Sheep Cooperative Research Centre, Armidale, NSW, 2351

SUMMARY

Yearling fibre diameter profiles from the OFDA2000 instrument were used to derive records on mean and coefficient of variation (CV) of fibre diameter at multiple points in the post-weaning age window, from 20 to 90% of staple profile length. Genetic correlations were calculated between these traits and their yearling equivalents. Results showed that from 50% of the staple profile and higher, post-weaning fibre diameter is genetically the same trait as yearling fibre diameter. Although the derivation of CV of fibre diameter was less accurate, genetic correlations with the yearling expression were greater than 0.9 for all except one percentile point. The expected correlation between post-weaning and yearling fleece weight was also derived, and under simple assumptions it is not unreasonable to expect estimates of genetic correlations greater than 0.9.

INTRODUCTION

The MERINOSELECT genetic evaluation system provides Australian Sheep Breeding Values (ASBVs) for wool traits at three age stages: yearling (approximately 12 months of age), hogget (18 months), and adult (2 years and older). Expressions of a particular trait, fibre diameter for example, are treated as separate correlated traits at each age stage. For efficient selection and marketing of rams, ram breeders would like to accurately evaluate wool traits as soon as records can viably be collected. Consequently, there is considerable interest in adding post-weaning wool traits to MERINOSELECT, recorded prior to yearling and potentially from 6 months of age. In order to implement these traits as ASBVs it is necessary to estimate genetic parameters including correlations with traits at later stages. While post-weaning wool records have been accumulating over recent years, including in designed research trials, the difficult combination to estimate is between post-weaning and yearling because there is very limited scope to shear animals twice between 6 and 12 months of age. For fibre diameter however, it is possible to derive records in the approximate window of post-weaning ages, from a yearling record measured using the OFDA2000 instrument (Baxter 2001) which records the fibre diameter distribution in 5mm increments along the wool staple. This study presents correlations between post-weaning mean and CV of fibre diameter derived from OFDA2000 profiles, with mean and CV of fibre diameter, clean fleece weight, and staple strength measured at the yearling stage. The expected correlation between post-weaning and yearling fleece weight was also derived.

MATERIALS AND METHODS

Data. OFDA2000 fibre diameter profiles recorded on yearling fleeces were obtained from Merino progeny born in the Sheep CRC Information Nucleus (IN) flocks between 2007 and 2011 (van der Werf *et al.* 2010). Using profiles on individual animals where the OFDA2000 staple length was greater than 50mm, mean fibre diameter was calculated as the weighted average fibre diameter from the tip of the staple (i.e. the start of the wool growth period at birth) to points extending from 20% to 90% of the staple in 10% steps. That is, a weighted average of the mean diameters at each 5mm increment was calculated within each growth period percentile range, using number of fibres counted as the weights. Each percentile range was then treated as a trait,

* AGBU is a joint venture of NSW Dept. of Primary Industry and the University of New England

with the range of traits encompassing the post-weaning wool growth period. It is important to note that across animals, variation in wool growth rates mean that at any one percentile point there will be an unknown range in the underlying measurement date. This is a potential source of inaccuracy, although age adjustments are generally not made in analyses of fibre diameter.

CV of fibre diameter was derived for the same percentile ranges, but in order to account for variation both along and across fibres, 50,000 random deviates were sampled from a normal distribution defined by the observed mean and standard deviation of fibre diameter at each 5mm increment, before combining across all 5mm increments within a percentile range to estimate the overall CV of diameter for the percentile range. This is a further potential source of inaccuracy.

Correlations were estimated between these traits and yearling mean fibre diameter (yfd) and yearling CV of fibre diameter measured using the Laserscan instrument, based on an independent sub-sample of wool from the same fleece. Two other key breeding objective correlations were also investigated, between mean fibre diameter at each percentile point and yearling clean fleece weight (ycfw), and between CV of fibre diameter at each percentile point and yearling staple strength (yss). A summary of the data structure by trait is shown in Table 1.

Analyses. Bivariate animal models were fitted in ASReml (Gilmour *et al.* 2009), using the procedure developed by (Swan *et al.* 2015) for data from the IN flocks. Briefly, in addition to standard fixed effects, random effects for genetic groups defined either by flock of origin or Merino strain, additive direct genetic effects, and sire by flock interactions were fitted for each trait. Correlated variance structures were fitted for genetic group and additive direct effects, but not for sire by flock. A random maternal permanent environment effect was also fitted for ycfw. Estimates of genetic correlations were derived from the additive direct genetic effect only.

The expected correlation between post-weaning and yearling fleece weight. There is a part-whole relationship between post-weaning and yearling fleece weights due to the fact that if an animal is shorn as a yearling, all of the wool that was present when the animal was at post-weaning age is contained in the yearling fleece. Yearling fleece weight (Y) can therefore be expressed as $Y = P + (Y - P) = P + D$ where P and D represent post-weaning fleece weight and the weight of wool grown between a post-weaning time point and yearling shearing respectively. The correlation between post-weaning and yearling fleece weight (r_{PY}) can then be derived as $r_{PY} = (r_{PD}\sigma_D + \sigma_P)/\sigma_Y$. If the correlation between P and D is zero this simplifies to $r_{PY} = \sigma_P/\sigma_Y$, the square root of the variance ratio between post-weaning and yearling fleece weight. For example, for a variance ratio of 50%, the expected correlation is 0.71. More likely is that P and D are correlated to a similar degree as two independent fleece weights, say yearling and adult. With fixed values of σ_P and σ_Y and an assumed value of r_{PD}, the value of σ_D can be derived by numerical optimisation, and the expected correlation r_{PY} can be calculated at the phenotypic and genetic levels.

Table 1. Data summary for traits analysed including OFDA2000 mean and CV of fibre diameter (yfd_o and ydcv_o), Laserscan mean and CV of fibre diameter (yfd and ydcv), clean fleece weight (ycfw) and staple strength (yss).

Statistic	yfd_o	ydcv_o	yfd	ydcv	ycfw	yss
Units	micron	%	micron	%	Kg	N/Ktex
Records	4948	4948	5137	5113	5683	4872
Trait mean	17.3	18.3	16.8	18.6	2.4	31.7
Trait SD	1.9	2.0	1.6	2.7	0.7	11.8
Mean age	314	314	330	330	336	331
Sires	200	200	186	186	210	186
Dams	3086	3086	3279	3275	3570	3167
Genetic groups	126	126	126	126	126	126

RESULTS AND DISCUSSION

Results in Table 2 show that for as little as 20% of the yearling staple, the heritability of fibre diameter exceeded 0.5, and the genetic correlation with yearling diameter was close to 0.9 or higher. At 50% of the staple, the genetic correlation reached 0.97, and at 70% of the stable, 0.99. Heritability also increased, to 0.74 at 90% of the staple. This demonstrates that at the likely ages post-weaning wool traits will first be included in MERINOSELECT in the window between 50 and 90% of the staple, mean fibre diameter is genetically the same trait as yearling fibre diameter. Further, it is apparent that mean fibre diameter can be accurately measured for the purpose of genetic analyses from even younger ages. These high heritabilities and genetic correlations were observed even under the circumstances that the measurement points on the OFDA2000 profiles could not be matched to consistent measurement dates across animals, as described above.

The genetic correlations between fibre diameter along the staple and yearling clean fleece weight (Table 2) showed a declining trend, from 0.48 at 20% of the staple to 0.31 at 90%, with the latter effectively the same as the estimate of 0.32 between yfd and ycfw reported by Swan *et al.* (2015) from the same data. Higher genetic correlations between fleece weight and fibre diameter measured at early ages are a cause for concern, and warrant further investigation. An initial step would be to see if the pattern is repeated in OFDA2000 profiles from the second shearing of these animals.

Table 2. Parameter estimates for mean fibre diameter (± standard error) derived from OFDA2000 staple profiles with varying amounts of the staple included (% of staple), including heritability, genetic and phenotypic correlation with Laserscan yearling fibre diameter (r_g yfd and r_p yfd) and yearling clean fleece weight (r_g ycfw and r_p ycfw).

% of staple	Heritability	r_g yfd	r_p yfd	r_g ycfw	r_p ycfw
20	0.54 ± 0.05	0.88 ± 0.02	0.67 ± 0.01	0.48 ± 0.10	0.21 ± 0.02
30	0.56 ± 0.05	0.91 ± 0.02	0.71 ± 0.01	0.46 ± 0.10	0.22 ± 0.02
40	0.58 ± 0.05	0.94 ± 0.01	0.75 ± 0.01	0.44 ± 0.10	0.21 ± 0.02
50	0.62 ± 0.05	0.97 ± 0.01	0.79 ± 0.01	0.41 ± 0.10	0.21 ± 0.02
60	0.65 ± 0.05	0.98 ± 0.01	0.82 ± 0.01	0.39 ± 0.10	0.20 ± 0.02
70	0.66 ± 0.05	0.99 ± 0.01	0.85 ± 0.01	0.36 ± 0.10	0.20 ± 0.02
80	0.69 ± 0.05	1.00 ± 0.01	0.87 ± 0.00	0.35 ± 0.10	0.20 ± 0.02
90	0.74 ± 0.05	0.99 ± 0.00	0.90 ± 0.00	0.31 ± 0.10	0.19 ± 0.02

Results in Table 3 show high estimates of genetic correlations between derived CV of diameter and Laserscan yearling CV of diameter (0.88 to 0.97), although these were lower than the equivalent estimates for mean diameter in Table 2 and did not show the same pattern of increase with staple percentile. Genetic correlations were highest between 30 and 50% of the staple (0.95 to 0.97). By contrast, heritability of the derived measurements did increase, from 0.19 at 20% to 0.31 at 90%, although this was still considerably lower than the estimate of 0.50 for Laserscan ydcv reported by Swan *et al.* (2015) from the same data. The genetic correlation between derived CV of diameter and yearling staple strength followed the expected trend, ranging from -0.50 to -0.61. Although it appears highly likely that the method used to derive CV of diameter for partial staples in these data has reduced accuracy, indications are that the trait can be measured during the post-weaning period and will be a useful selection criterion for staple strength. These results are consistent with the findings of Greeff and Schlink (2013) that CV of diameter on partial staples was genetically the same trait as CV on the whole staple. Interestingly, Greeff and Schlink also found that staple strength could be accurately assessed on staples from 60% of the staple length and higher (genetic correlation estimates of 0.85 for 60%, and 0.99 for 80% of the staple).

Expected correlations between post-weaning and yearling clean fleece weight (pcfw and ycfw) are shown in Table 4. The phenotypic variance of 0.16 for ycfw was taken from Swan *et al.*

(2015), while the value 0.08 for pcfw was from unpublished MERINOSELECT data (Brown, 2015, pers. comm.). To derive genetic variances, heritability was assumed to be 0.3 at both stages, and genetic and phenotypic correlations between P and D were set to the equivalent correlations between yearling and adult clean fleece weight used in MERINOSELECT. Based on these input values, the expected genetic and phenotypic correlations were 0.96 and 0.92 respectively. While yet to be confirmed by estimates from industry data, these theoretical results strongly suggest that fleece weight recorded during the post-weaning stage window will provide accurate information on yearling fleece weight. Collection of fleece weights in industry flocks early in the post-weaning window (around 6 months of age) would be highly desirable to determine how early fleece weight can be measured. The loss in accuracy due to measuring at young ages could potentially be offset by the ability to shorten the generation interval of males leading to increased rates of genetic gain.

Table 3. Parameter estimates for CV of fibre diameter (± standard error) derived from OFDA2000 staple profiles with varying amounts of the staple included (% of staple), including heritability, genetic and phenotypic correlation with Laserscan yearling CV of fibre diameter (r_g ydcv and r_p ydcv) and yearling staple strength (r_g yss and r_p yss).

% of staple	Heritability	r_g ydcv	r_p ydcv	r_g yss	r_p yss
20	0.19 ± 0.04	0.93 ± 0.06	0.39 ± 0.01	-0.50 ± 0.11	-0.16 ± 0.02
30	0.23 ± 0.04	0.95 ± 0.05	0.41 ± 0.01	-0.54 ± 0.10	-0.19 ± 0.02
40	0.25 ± 0.04	0.97 ± 0.05	0.43 ± 0.01	-0.57 ± 0.09	-0.21 ± 0.02
50	0.27 ± 0.04	0.97 ± 0.05	0.45 ± 0.01	-0.58 ± 0.09	-0.24 ± 0.02
60	0.28 ± 0.04	0.92 ± 0.04	0.48 ± 0.01	-0.61 ± 0.08	-0.27 ± 0.02
70	0.29 ± 0.04	0.90 ± 0.04	0.50 ± 0.01	-0.60 ± 0.08	-0.29 ± 0.02
80	0.29 ± 0.04	0.90 ± 0.04	0.51 ± 0.01	-0.61 ± 0.08	-0.30 ± 0.02
90	0.31 ± 0.05	0.88 ± 0.04	0.53 ± 0.01	-0.60 ± 0.08	-0.30 ± 0.02

Table 4. Expected correlations between post-weaning and yearling clean fleece weight (r_{PY}) assuming correlations between P and D (r_{PD}) match MERINOSELECT correlations between yearling and adult fleece weight, and observed variances for post-weaning and yearling fleece weight (σ_P^2 and σ_Y^2).

Correlation	σ_P^2	σ_Y^2	r_{PD}	r_{PY}
Genetic	0.024	0.048	0.71	0.96
Phenotypic	0.08	0.16	0.48	0.92

CONCLUSIONS

The results presented in this study demonstrate that Merino breeders can have confidence in the accuracy of ASBVs for post-weaning wool traits as selection criteria for yearling wool breeding objective traits. Further research to identify how early fleece weight can be recorded in the post-weaning age window, and whether breeding programs using early wool measurements can lead to increased rates of genetic gain would be of value.

ACKNOWLEDGEMENTS

This work was partly funded by MLA Project B.SGN.0127.

REFERENCES

Baxter, B.P. (2001) *Wool Technology and Sheep Breeding* **49**: (1)
Gilmour, A.R. *et al.* (2009) *ASReml User Guide Release 3.0*. Hemel Hempstead, HP1 1ES, UK.
Greeff, J.C., and A.C. Schlink. (2013) *Proc. Conf. Assoc. Adv. Anim. Breed. Genet.*, **20**:393.
Swan, A.A., D.J. Brown, and J.H.J. van der Werf. (2015) *Anim. Prod. Sci.* http://dx.doi.org/10.1071/AN14560.
Van der Werf, J., B. Kinghorn, and R. Banks (2010) *Anim. Prod. Sci.* **50**:998.

MEASUREMENT OF METHANE TRAITS IN THE BEEF INFORMATION NUCLEUS CATTLE

P.F. Arthur[1], R.M. Herd[2], C. Weber[2], H. Smith[2] and R.F. Hegarty[3]

[1] NSW Department of Primary Industries, Agricultural Institute, Menangle, NSW 2568
[2] NSW Department of Primary Industries, Beef Industry Centre, Armidale, NSW 2351
[3] Environmental and Rural Science, University of New England, Armidale, NSW 2351

SUMMARY

Methane is the main source of greenhouse gas (GHG) from ruminants. The focus of this study was to measure methane production (MPR) in industry cattle. The ultimate aim was to develop estimated breeding values (EBVs) for methane traits for use in genetic improvement to reduce GHG in cattle. A total of 548 Australian Angus and 102 Australian Charolais Beef Information Nucleus herd steers which were undertaking a net feed intake test at the University of New England "Tullimba" research feedlot, near Armidale NSW, were measured for MPR using GreenFeed Emission Monitoring (GEM) units. The units provide short term MPR measurement anytime an animal visits a unit. Two-thirds of the steers from each breed visited the GEM units; most of them more than 20 times within a period which ranged from 50 to 66 days. There were significant ($P < 0.05$) sire differences in MPR, MPR per unit feed intake, MPR per unit body weight and MPR per unit average daily gain in Angus cattle. The significant sire differences in the methane traits indicate the presence of genetic variation for methane traits when measured by GEM units.

INTRODUCTION

Across the globe the agricultural sector is a significant source of GHG emissions. A recent Food and Agriculture Organisation (FAO) study reported that livestock were responsible for 14.5% of global GHG emissions, and ruminants contribute about 80% of the livestock emissions (Gerber *et al*. 2013). Methane is the main source of GHG from ruminants, and is produced during the process of microbial fermentation of plant material, mainly in the rumen, which is then exhaled into the atmosphere. High methane production is associated with high feed intake in ruminants (Blaxter and Clapperton 1965; Pelchen and Peters 1998). Feed intake is highly correlated with growth and other productivity traits in ruminants (Arthur *et al*. 2001), hence any strategy to lower methane production *per se,* may have a detrimental impact on ruminant productivity through a correlated reduction in feed intake. Hence there has been increased interest in the amount of methane produced per unit feed intake, also known as methane yield (MY) and the amount of methane produced per unit of product, known as methane intensity (MI).

In the last decade there has been active development of methane measurement technologies to the stage where individual animal methane production (MPR) can now be measured on a large scale. The suitability and accuracies of these methane measurement technologies are currently being assessed, in conjunction with the development of standardised measurement protocols for livestock. For genetic improvement purposes it is essential to accurately measure the trait of interest (e.g. methane) in a large number of animals on a regular basis to estimate the genetic merit of each potential breeding animal for the trait. The focus of this study was to measure methane in large numbers of industry cattle from high-profile potentially elite sires that will contribute to future genetic improvement in Australian cattle breeds.. The ultimate aim was to develop estimated breeding values for methane traits for use in genetic improvement.

MATERIALS AND METHODS

Experimental animals. The Australian beef industry has developed a number of resource herds known as the Beef Information Nucleus (BIN) to underpin research and development of genomic selection. The BIN is a progeny testing scheme (Banks, 2011) and current participants include the Angus, Brahman, Charolais, Hereford and Limousin breed societies. Records on all economically important traits are collected on steers, heifers and cows, including difficult to measure traits like feed intake and feed efficiency. The steers from the Angus and Charolais BINs are measured for growth and net feed intake (NFI; the measure of feed efficiency) on a feedlot ration at the University of New England "Tullimba" research feedlot, near Armidale NSW. While being measured for NFI, the MPR of some cohorts of steers were measured. From December 2013 to January 2015, one cohort of Charolais and five cohorts of Angus BIN cattle had the opportunity to be measured for MPR. Each cohort is made up of approximately equal numbers of steers per sire used that in the progeny test for the year.

Measurement technology. The MPR of the cattle was measured using GreenFeed Emission Monitoring (GEM) units manufactured by C-Lock Inc. (U.S. Patent 7966971). The GEM units provide short term measurement of methane production made many times per day using portable breath collection and methane analyzer. The scientific principles underpinning the operation of the unit is explained by Zimmerman *et al*. (2013), and a detailed description is provided by Zimmerman (2013) at the company's website (http://c-lockinc.com/whatisgreenfeed.php). At the research feedlot the GEM units were located next to the feed intake recorders and the steers have the opportunity to voluntarily visit the GEM unit at any time to have their MPR measured.

Statistical analysis. The definitions of all the traits used in this report are provided in Table 1. The data used were from 102 Charolais (1 cohort) and 548 Angus (5 cohorts) steers who had the opportunity to visit the GEM units during their NFI test. The percentage and frequency of visits to the GEM units were calculated for both breeds. The number of records and progeny per sire in the Charolais data was limited and thus was not subjected to further analysis. The data from the Angus steers were analysed further to assess sire differences for the methane traits using records on the 356 steers with a minimum of 20 visits to the GEM units. A preliminary analysis was conducted to examine cohort effect on the traits with a full model which included the effects of cohort, sire, sire by cohort and residual error. All the terms in the model were assumed to have random effects and the errors were assumed to have a normal distribution with mean zero and variance. Cohort and sire effects were highly confounded hence cohort and sire by cohort were dropped in the final analysis. All parameters were estimated using residual maximum likelihood (REML) procedure and the sire means were adjusted using the best linear unbiased predictor (BLUP) due to their random nature. Least significant difference at 5% level was calculated and used to compare means between sires. All analyses were run on ASReml Release 3.00 (Gilmour *et al*, 2009).

Table 1. Definition of traits

Trait name	Abbreviation	Units	Definition
Body weight	BW	kg	Mid test BW (Start BW + End BW)/2
Average daily gain	ADG	kg	Daily BW gain (End BW – Start BW)/ days on test
Daily feed intake	FI	g/day	Average daily feed intake during the test
Methane production rate	MPR	g/day	Methane produced
Methane yield	MY	g/kg	MPR per unit FI (MPR ÷ FI)
Methane intensity$_{BW}$	MI_B	g/kg	MPR per unit BW (MPR ÷ BW)
Methane intensity$_{ADG}$	MI_A	g/kg	MPR per unit ADG (MPR ÷ ADG)

RESULTS AND DISCUSSION

About one third of the steers (33% of Angus and 34% of Charolais) did not visit the GEM units, hence did not have an MPR record. The percentage of steers with more than 20 visits to the GEM units was 56% for Angus and 63% for Charolais. Therefore, in order to obtained accurate data on as many progeny per sire as possible for the estimation of genetic parameters and EBVs for methane traits, the number of steers which use the GEM units and the frequency of visits need to increase. Measurement of MPR with the GEM units relies on the cattle visiting the units. In the feedlot environment the steers had *ad libitum* access to the high-energy feedlot ration. Although the GEM unit deliver a small amount of feed pellets to the steers on each visit, it did not appear to be a strong enough attractant for many steers. It is therefore important that other inducement strategies be investigated, especially in the feedlot environment. Most of the steers that used the GEM units visited more than 20 times during their NFI test period which were from 50 to 66 days.

The 356 Angus steers who had frequent visits (>20) to the GEM units were the progeny of 83 sires, with an average of 4.3, and a range of 1 to 14 steers per sire. The descriptive statistics for all the traits studied in the Angus steers are presented in Table 2. The steers were consuming an average of 15.8 kg of feed, growing at 1.95 kg and producing 217.6 g of methane per day.

Table 2. Descriptive statistics for production and methane traits of Angus steers

Trait[1]	Mean	SE	Min	Max
Body weight (BW), kg	601.9	3.7	438	830
Average daily gain (ADG), kg	1.95	0.02	0.72	3.53
Daily Feed intake (FI), kg	15.8	0.11	10.5	22.7
Methane Production rate, g/day	217.6	2.05	104.5	385
Methane yield, g/kg FI	13.86	0.14	6.41	26.93
Methane intensity$_{ADG}$, g/kg ADG	115.3	1.5	58.0	243.7
Methane intensity$_{BW}$, g/kg BW	0.363	0.003	0.166	0.640

[1]See Table 1 for full trait names and definitions

There were significant ($P < 0.05$) sire differences in all the methane traits (MPR, MY, MI_B and MI_A). Means for MY and MI_A of the top and bottom five sires are presented in Figure 1. In general, the sire differences were significant only between the top and bottom five sires for each of the methane traits, with no significant differences among the sires in the middle range. Donoghue *et al*. (2015) has shown that there is genetic variation, with moderate heritability for methane traits in Angus cattle when measured in respiration chambers. The presence of significant sire differences in the methane traits of the steers in the feedlot indicates the presence of genetic variation for methane traits when measured by GEM units. However, the number of animals measured is not adequate for comprehensive analyses to provide an accurate estimate of the genetic variation for these methane traits when measured in the GEM units. It is therefore essential that measurement of animals in the GEM units be continued in order to generate enough GEM based data for the estimation of genetic parameters and for the development of EBVs for industry cattle. The research on the measurement of methane traits on females at pasture is in its infancy and the use of GEM units to collect data from females at pasture needs to continue so that the relationship between feedlot measurements in steers and pasture measurement in females can be evaluated. Very few animals have been measured on pasture and measurement protocols are currently being developed and trialled. Continued funding is required to progress the work on measuring methane on females at pasture.

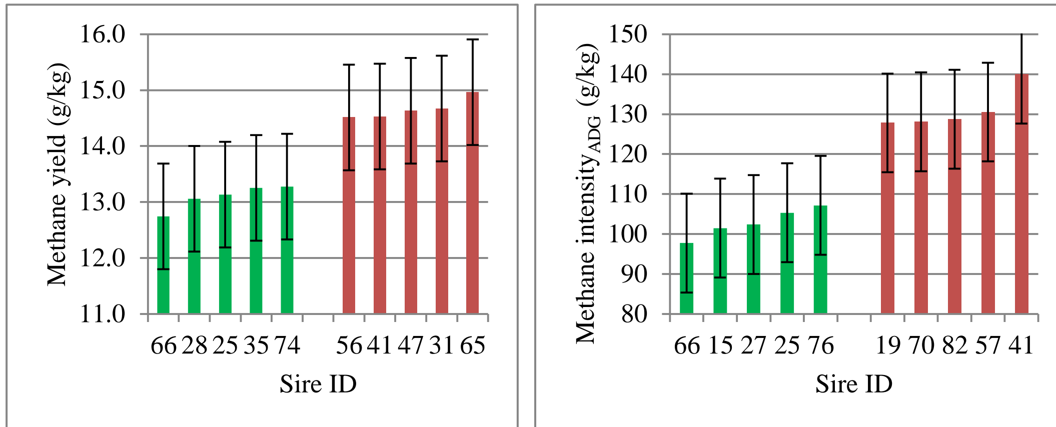


Figure 1. Mean methane yield and methane intensity$_{ADG}$, with standard error bars of steer progeny for the top (red) and bottom (green) five Angus sires

ACKNOWLEDGEMENTS

Financial support was provided by Meat and Livestock Australia and the Australian government Department of Agriculture, as part of the National Livestock Methane Program. Assistance provided by Idris Barchia, Jim Cook, Simon Bird and Reg Woodgate is acknowledged.

REFERENCES

Arthur P.F., Archer J.A., Johnston D.J., Herd R.M., Richardson E.C. and Parnell P.F.. (2001) *J. Anim. Sci.* **79**: 2805.

Banks R.G. (2011) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **19**: 399.

Blaxter K.L. and Clapperton J.L. (1965) *Brit. J. Nutr.* **19**: 511.

Donoghue K.A., Bird-Gardiner T., Arthur P.F., Herd R.M. and Hegarty R.F. (2015) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21**: (In Press).

Gerber P.J., Steinfeld H., Henderson B., Mottet A., Opio C., Dijkman J., Falcucci A. and Tempio G. (2013) Tackling climate change A. through livestock – A global assessment of emissions and mitigation opportunities. FAO, Rome.

Gilmour A.R., Cullis B.R., Welham S.J. and Thompson R. (2009) 'ASREML: Program user manual' NSW Agriculture, Orange Agricultural Institute, Forest Road, Orange, NSW, 2800, Australia.

Pelchen A. and Peters K.J. (1998) *Small Rumin. Res.* **27**: 137

Zimmerman S. 2013. What is GreenFeed? http://c-lockinc.com/whatisgreenfeed.php

Zimmerman S., Michal J.J., White R., Johnson K.A., Guerouali A. and Zimmerman P. (2013) *Proc. Am. Dairy Sci. and Am. Soc. Anim. Sci.* pp. 527.

REGIONS IMPACTING INBREEDING DEPRESSION AND THEIR ASSOCIATION WITH ADDITIVE GENETIC EFFECTS FOR JERSEY CATTLE FROM THE UNITED STATES OF AMERICA AND AUSTRALIA

J. T. Howard[1], J. E. Pryce[2,3], M. Haile-Mariam[2], C. Maltecca[1]

[1]Department of Animal Science, North Carolina State University, Raleigh, NC 27695-7627, USA
[2]Department of Economic Development, Jobs, Transport and Resources and Dairy Futures Cooperative Research Centre, 5 Ring Road, Bundoora Victoria, 3083, Australia
[3]La Trobe University, Bundoora, Victoria 3086, Australia

SUMMARY

The advent of dense genotype platforms allow for the identification of specific regions that give rise to inbreeding depression and characterize their relationship with the additive effect of that region. Utilizing a run of homozygosity (ROH) metric the first study objective is to identify regions having an impact on inbreeding depression for United States (US) and Australia (AU) Jersey cows. The second objective is to determine the relationship between the additive and ROH SNP effects. Genotyped cows with phenotypes on milk yield traits (US=6751; AU=3974) and calving interval (US=5816; AU=3905) were utilized. A ROH based metric (ROH4Mb) was calculated across the genome. Residuals from a model that accounted for the fixed and additive genetic effects were regressed on ROH4Mb using a single marker regression or a machine-learning tree based model. The relationship between ROH4Mb and additive effect was characterized based on sliding window (500kb) direct genomic value derived from a Bayesian-LASSO analysis. Genomic regions across multiple traits were found to be associated with ROH4Mb for the US on BTA13, BTA23 and BTA25 and AU on BTA3, BTA7 and BTA17. Multiple potential epistatic interactions were characterized. The covariance between ROH4Mb and the additive effect depended on the genomic region.

INTRODUCTION

High levels of inbreeding result in a reduction in fitness and overall performance at the phenotypic level (Leroy, 2014). Previous research has shown that inbreeding depression is heterogeneous across founders (Gulisija *et al.* 2006), which implies that the genetic load is not distributed evenly among genomes. Utilizing a run of homozygosity (ROH) metric, Pryce *et al.* (2014) confirmed heterogeneity in inbreeding depression by identifying multiple regions that resulted in reduction in milk yield and fertility traits in Australian (AU) Holstein and Jersey cows. Furthermore, regions with multiplicative effects, which individually may have a minor effect but which have significant impact on fitness when combined, might provide clues about the previously identified non-linear relationship of inbreeding depression (Gulisija *et al.* 2007).

Characterizing regions that give rise to inbreeding depression in dairy cattle is advantageous due the increasingly large number of genotyped cows and the extensive list of recorded phenotypes. It has been shown by Howard *et al.* (2015) that ROH frequency differs across the Australian (AU) and United States (US) Jersey populations, which could potentially give rise to different regions that have an impact on inbreeding depression. Also, Howard *et al.* (2015) found that regions of the genome with high ROH frequency are most likely the result of directional selection. Here we hypothesize that the covariance between the additive effect and ROH status of a SNP is variable across the genome, and characterizing this may provide clues about the relationship between the two metrics. Therefore the

objectives are: 1. Identify regions that have an impact on inbreeding depression for US and AU Jersey cows utilizing ROH metrics; 2. Determine the relationship between the additive and ROH status effect of a SNP.

MATERIALS AND METHODS

Data. Two populations of genotyped cows born in the US and AU were utilized to identify regions that when homozygous cause a reduction in dairy yield traits including milk (MY), fat (FY) and protein (PY; n = 6751 US; n = 3974 AU) and fertility measured as calving interval (FERT; n = 5816 US; n = 3905 AU). Phenotypic information for the AU population was provided in the form of yield deviations. In order to make comparison similar, yield deviations were also constructed for the US population based on the same model outlined by Howard *et al.* (2015). A total of 32,431 QC SNP were used for the analysis. The ROH status of a SNP was defined based on whether the SNP was within an ROH of at least 4 Mb in length (ROH4Mb; 1 if the SNP was in a ROH and 0 otherwise).

Genome wide association study (GWAS). A two-stage analysis was performed within each population. Stage one involved generating residuals from an animal model that accounted for the additive genetic effects. The second stage involved using the residuals from the first stage as a phenotype to regress on ROH4Mb status. For this analysis two models were utilized. Model 1 was a single marker regression model, which regressed the phenotype on the ROH4Mb status for each SNP. Significance (P-value < 0.001) was declared by using a permutation test (n= 2,500 samples). The second model utilized a machine-learning tree based regression algorithm, referred to as a gradient boosted machine (GBM), to identify higher order ROH interactions. The "gbm" R package (Ridgeway 2010) was used to carry out the analysis. Based on a 4-fold cross-validation the final model was constructed from 1200 trees at an interaction depth of 5 and a shrinkage parameter of 0.0075. Within each chromosome, SNP with a correlation exceeding 0.1, as outlined by Lubke *et al.* (2013), were removed and only SNP with the largest impact based on single marker regression analysis were included in the final analysis. The final number of SNP utilized was 115 and 81 for the US dataset and 100 and 105 for the AU dataset for yield traits and fertility, respectively. The variable importance value (Ridgeway 2010) was used as measure to assess the importance of a ROH4Mb status of a SNP on a given phenotype.

The identification of epistatic interactions between the ROH4Mb status of a SNP was carried out by counting the number of times two SNP were a descendent pair, as described by Yao *et al.* (2013). Briefly, based on Figure 1, assume SNP B and D have a large epistatic interaction. The SNP pairs are represented based on the levels at which they appear, such that SNP D was derived from a split produced by SNP B, and therefore the combination represents a parent (SNP B) and child (SNP D) descendent pair. The SNPs B and D will appear more frequently in the same branch of a tree due to the pair having an epistatic interaction. Lower level descendent pairs such as parent (SNP A) grandchild (SNP D) will also be referred to as a descendent pair. The identification of SNP with independent effects, such as SNP B and C won't be tagged as descendent pairs due to SNP B and C being on separate branches. The significance (P-value < 0.001) of the frequency of a descendent pair and variable importance value was declared based on a permutation test (n= 2,500 samples).

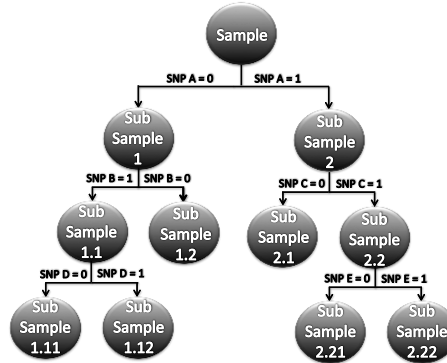


Figure 1. An example of a regression tree generated by a Gradient Boosted Machine algorithm.

Additive and ROH4Mb relationship. The second objective of the study was to characterize the relationship between the additive and ROH4Mb effect of a SNP. Both the additive and ROH4Mb marker values were obtained within each population and trait using a Bayesian LASSO marker regression (Park and Casella 2008). Estimates of the additive effect of a SNP were obtained using yield deviations as phenotypes. Estimates of the ROH4Mb effect of a SNP were obtained using the same phenotype as in the two-stage approach. The LASSO analysis was performed using the 'BLR' package in R (de los Campos *et al.* 2013). To characterize the relationship between the additive genetic and ROH4Mb SNP effect across the genome, 500 kb overlapping windows were used to estimate the GEBV (co)variance for a given window for both analyses.

RESULTS AND DISCUSSION

Within a population, regions that had an effect across multiple traits included BTA13 (19.3-19.9 Mb; MY-PY), BTA23 (32.7-33.3; MY-FY-PY) and BTA25 (24.8-30.7; MY-PY) for the US population and BTA3 (113.4-114.6; FY-PY), BTA7 (6.6-16.7; FY-PY) and BTA17 (68.9-75.0; MY-FY-PY) for the AU population. In both countries none of these regions had an effect on calving interval. Strikingly, no regions were identified that were significant in both populations. Both models ranked regions comparably with a rank correlation of 0.48 to 0.65 across all traits and populations. A list of the regions and candidate genes is outlined in Table 1.

Multiple regions of the genome were found to display potential interactions based on their frequency of being a descendent pair. The majority of all significant descendent pairs were associated with at least one SNP that also had a large variable importance score. Additionally, a gene network analysis revealed network associations including shared protein domain (HNF1B-LBX2) genetic interactions (HNF1B-LOXL3; HNF1B-RPL17). Additional work will be needed to validate regions found across both the single marker regression and GBM analyses in other populations.

The relationship between the additive genetic effect and the ROH4Mb effect displayed positive and negative covariance across the genome. Regions on BTA3, BTA7, BTA20 and BTA26 had a positive covariance between the additive and ROH4Mb effect of a SNP across both populations, and

have previously been found to be under positive directional selection (Howard *et al.* 2015). Homozygosity at certain regions of the genome is beneficial such that homozygosity based on the ROH4Mb status gives rise to a higher additive genomic estimated breeding value. Conversely, homozygosity at certain regions is unfavorable giving rise to lower additive genomic estimated breeding values. The majority of the regions with the largest absolute covariance value across traits were positive, which is not surprising due to a low frequency of ROH4Mb status for regions with a large ROH4Mb effect (mean ROH4Mb frequency = 0.089) in comparison to the regions that displayed a large positive covariance (mean ROH4Mb frequency = 0.235). These results provide evidence that it is possible to distinguish between two individuals that have the same inbreeding coefficient, but different overall genomic loads. This would have in turn important consequences in the management of genomic diversity and the implementations of effective mating design.

Table 1. Regions of the genome associated with inbreeding depression for milk traits.

Country	Traits	BTA (Region Mb)	Frequency	Candidate Gene
United States	MY-PY	13 (19.3-19.9)	0.10	PARD3
	MY-FY-PY	23 (32.7-33.3)	0.18	ALDH5A1
	MY-PY	25 (24.8-30.7)	0.05	IL4R, CALN1
Australia	FY-PY	3 (113.4-114.6)	0.06	UGT1A1
	FY-PY	7 (6.6-16.7)	0.17	NOTCH3
	MY-FY-PY	17 (68.9-75.0)	0.04	IGLL1

REFERENCES

de Los Campos G., Pérez P., Vazquez A.I. and Crossa J. (2013) *Methods Mol Biol* **1019**: 299.

Gulisija, D., Gianola D. and Weigel K.A. (2006) *Livest. Sci.* **104**: 244.

Gulisija, D., Gianola D. and Weigel K.A. (2007) *J. Dairy Sci.* **90**: 493.

Howard J.T., Maltecca C., Haile-Mariam M., Hayes B.J. and Pryce J.E. (2014) *BMC Genomics* **16(1)**: 1352.

Leroy, G. (2014) *Anim. Genet.* **45**: 618.

Lubke G.H., Laurin C., Walters R., Eriksson N., Hysi P., Spector T.D., Montgomery G.W., Martin N.G., Medland S.E. and Boomsma D.I. (2013*) J Data Mining Genomics Proteomics* **4**:10.4172/2153-0602.1000143.

Park T. and Casella G. (2008) *J. Am. Statist. Assoc.* **103**: 681.

Pryce, JE, Haile-Mariam M., Goddard M.E. and Hayes B.J. (2014) *Genet. Sel. Evol.* **46**: 71.

Ridgeway G. (2010) *R package version* 1.6-3.1.

Yao C., Spurlock D.M., Armentano L.E., Page Jr C.D., Van de Haar M.J., Bickhart D.M. and Weigel K.A. (2013) *J. Dairy Sci.* **96**:6716.

GENETIC AND PHENOTYPIC RELATIONSHIPS BETWEEN KID SURVIVAL AND BIRTH WEIGHT IN AUSTRALIAN MEAT GOATS

M.N. Aldridge[1], D.J. Brown[2] and W.S. Pitchford[1]

[1]School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy
[2] Animal Genetics and Breeding Unit*, University of New England, Armidale

SUMMARY

The Australian goat industry would like to improve reproductive rate by increasing kid survival. Parameter estimates for kid survival and correlated traits are yet to be reported. A preliminary analysis of birth weight and survival was conducted using 16,050 records from industry herds. The heritability for birth weight (0.32±0.029) was similar to previous reports, but the heritability for kid survival (0.29±0.024) was higher than expected in comparison to other breeds of goats and sheep. The phenotypic variance for birth weight is similar to those previously reported for Boer goats. For a binomial trait there was moderate variation in kid survival with a phenotypic deviation of 0.288, birth weight had a moderate amount of variation with a standard deviation of 0.599kg. The lowest kid survival rates occur in animals less than 2.5kg with survival rates between 67% and 85%, while animals over 2.5kg had survival rates between 92% and 98%, the overall mean for survival was 85%. The phenotypic correlation estimate of 0.16 is low but positive for birth weight and survival. The genetic correlation was also positive and high at 0.54±0.068. Improving survival could potentially be achieved either with direct selection or indirect selection with birth weight.

INTRODUCTION

The Australian goat industry was valued at $146m AUD in 2012/13 (MLA 2012). The majority of production is achieved by harvesting feral goats from the Australian rangelands but genetic improvement is focused on the domestic Boer goat population. While there is limited genetic research on goats in Australia, reproductive rate has been identified as a trait requiring improvement (MLA 2012). Goats are moderately fertile (does kidding/doe joined = 0.82, Walkden-Brown and Bocquier 2000) and highly fecund (kids born / doe joined = 1.76, Zhang *et al.* 2009). Instead of focusing on fertility and fecundity, increasing kid survival between birth and weaning could increase the net reproductive rate.

Australian sheep research has made recent progress on survival traits much of which is yet to be replicated in Australian goats. Sheep have been shown to display low to moderate genetic correlations between birth weight and survival but with curvilinear phenotypic relationships between these traits (Brien *et al.* 2014).

This paper reports a preliminary analysis using the KIDPLAN national genetic evaluation database to estimate the genetic and non-genetic relationships between kid survival and birth weight. This is part of a larger project which aims to determine whether correlated traits can be exploited to achieve genetic improvement of kid survival in Australian production systems.

MATERIALS AND METHODS

Data analysed for this study was submitted to the national performance recording scheme KIDPLAN and included records on 16,280 individuals born from 1991 to 2014. 973 duplicate records were removed during data cleaning. The animals are progeny of 574 sires and 3,669 dams. The median number of progeny per sire was nine and per dam was three. The key traits of interest

* AGBU is a joint venture of NSW Dept. of Primary Industry and the University of New England

were birth weight and kid survival. Birth weight was recorded by the breeder. Kid survival was generated using birth and rearing type records, 743 dead tags were added when multiple litter sizes did not have data submitted on the dead kid. Ideally breeders would submit data for all dead kids, from the calculated survival 2404 kids did not survive to weaning and of these 507 dead kids had a birth weight recorded. Due to the low number of quadruplets (480 kids) and quintuplets (60 kids), litters of 3 or greater were analysed as a single category termed 'multiples'. Fixed effects in the analysis were date of birth, sex, dam age and site in addition to birth type and rearing type. The data was submitted from 40 herds.

Means, phenotypic variances, heritabilities and correlations were estimated using ASREML-R (Butler *et al.* 2009) for both kid survival and birth weight. Univariate analysis of birth weight and kid survival was conducted with fixed effects for sex, site, year of birth, site by year of birth interaction, dam age and birth type included in the models. The animal model included the following random effects; direct genetic effect fitted as a trait of the kid, a permanent environmental and maternal genetic component and a sire by flock by year interaction. Predictions for birth weight and survival were also estimated with birth type as it had significant effect on both traits.

A bivariate analysis for birth weight and survival was conducted to estimate phenotypic variance, heritability, genetic and phenotypic correlations. The full model included the fixed effects birth type, sex, flock, year of birth, flock by year of birth interaction and dam age. The sire by flock by year interaction was removed from the random effects as it was not estimable in the bivariate model, leaving the random effects as direct genetic, permanent environmental and maternal genetic components.

RESULTS AND DISCUSSION

The mean birth weight (±SD) for goats in this data set was 3.5kg ±0.6 and was normally distributed. Kid survival rates for different birth weight categories are illustrated in Figure 1 and shows a trend for lower birth weights to be associated with lower survival rates. It also appears that high birth weight was not associated with lower survival as seen in other species due to dystocia (Brown *et al*. 2014). The survival rate for goats between 1.0 and 1.4kg was 67%, 1.5 to 1.9kg was 70% and goats 2 to 2.4kg was 85%. The mean survival rate for goats was 85%, which was higher than a mean Merino lamb survival rate of 72% from a sample with a similar mean birth weight of 3.63kg (Hatcher *et al*. 2014).

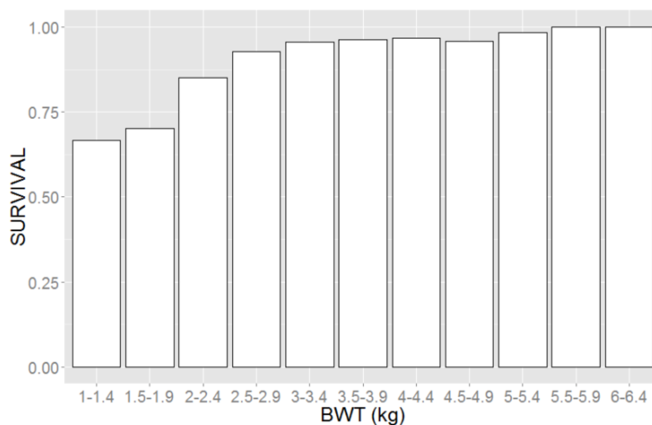


Figure 1. Survival rate of goats plotted against birth weight groups of 0.5kg

Predicted means for birth weight and kid survival rates for each birth type class were made using the univariate model. Birth type had a significant effect on both kid survival (P<0.01) and birth weight (P<0.01). Singles had the largest birth weight (3.88kg), twins (3.49kg) and multiples the lowest (3.16kg) (Table 1). There was no significant difference in kid survival between singles and twins (83% and 82% respectively) but multiples had a lower survival rate (71%).

Table 1. Predicted means (± se) of Boer kid survival rates for each birth type class and birth weight

Birth Type	Number of records for birth type	Survival	Birth Weight (kg)	Number of records for birth weight
1	3,510	0.83 (0.0075)	3.88 (0.019)	1,971
2	8,841	0.82 (0.0068)	3.49 (0.016)	6,301
≥ 3	3,699	0.71 (0.0084)	3.16 (0.020)	2,373

A summary of the heritability, phenotypic and genetic correlations, and phenotypic variances are summarised in Table 2.The heritability for birth weight was 0.32 and similar to other estimates for birth weight in Boer goats which have been reported between 0.19 and 0.39 (Ball *et al.* 2001, Zhang *et al.* 2008, Zhang *et al.* 2009). No studies are yet to publish heritability of survival in Boer goats however, the estimate of the current study of 0.29 was much larger than those reported for other breeds and species. Previous goat studies have reported heritabilities of 0.10 (Singh *et al.* 1990) and studies in sheep ranged between 0.01 and 0.03 (Brien *et al.* 2014). It is important to note that the phenotypic variance reported for birth weight is consistent with that of the Boer goat at 0.36 previously reported between 0.31 and 0.57 which allows for selection to be undertaken successfully (Ball *et al.* 2001, Zhang *et al.* 2008). Though the phenotypic variance of survival was low (0.08) there is variation in the trait, in comparison to sheep it is within the range of 0.05 and 0.19 as reported by Hatcher *et al*. (2014).

The genetic and phenotypic correlations between birth weight and kid survival are 0.54 and 0.16 respectively. The phenotypic correlation between birth weight and kid survival has previously been reported to be a curvilinear relationship (Snyman 2010). The genetic correlation has not yet been reported for goats and appears to be high. The genetic correlation for sheep has been reported between not different to zero and 0.45 (Brien *et al.* 2014).

Table 2. Estimates of, phenotypic variance (σ_p^2), permanent environmental variance (PE), maternal genetic variance (σ_m^2), heritability (h^2) and genetic (r_g) and phenotypic (r_p) correlations for birth weight and survival from a bivariate analysis of goats (±se).

Trait	Birth Weight(kg)	Survival
σ_p^2	0.36 (0.01)	0.08 (0.001)
PE	0.037 (0.01)	0.0034 (0.001)
σ_m^2	0.027 (0.01)	0.0008 (0.001)
h^2	**0.32 (0.03)**	**0.29 (0.02)**
Birth Weight	**-**	r_p = 0.16 (0.01)
Survival	r_g = 0.54 (0.07)	**-**

As the majority of losses occur in kids under 2.5kg, the results of this study suggest that selection to increase birth weight, to reduce mortalities, can be undertaken successfully though

caution must be taken not to develop dystocia problems. The preliminary results also suggest that improved survival could be achieved by direct selection for the trait.

The high heritability for survival was unexpected, and may be a consequence of the larger variation in birth type observed in the current study compared to sheep. Additionally the survival data is yet to be analysed using a more appropriate model for binomial data. The results could also be a function of the data if breeders did not submit complete kid mortality records.

The predictions for survival (Table 1) indicate that increasing survival could be accomplished by reducing litter size but this could have unfavourable consequences for weaning rate (as number of kids per doe joined). Future work will examine the relationship of birth weight with litter size and how selection to change these may influence weaning rates. Finally the results of this study suggest that kid survival in single, twin and multiple birth classes should be considered as separate traits and will also be investigated in future work.

CONCLUSIONS

The higher than expected heritability for kid survival reported in this study suggests that if measurements of kid survival are collected, and the trait is given priority in the breeding objective, breeders will be able to improve the trait by selection. The genetic and phenotypic correlation between kid survival and birth weight has implications for future indirect selection, but with a risk of increasing the chance of dystocia. The results also indicate that litter size could contribute to kid survival and birth weight which will be further investigated.

ACKNOWLEDGEMENT

This research was funded by Meat and Livestock Australia with gratitude to Sheep Genetics and the breeders contributing data to KIDPLAN.

REFERENCES

Ball, A., Brown, D., Spiker, S., Field, S., Banks, R. (2001) *Proc. Assoc. Adv. Anim. Breed. Gen* **14:** 445.

Brien, F, Cloete, S, Fogarty, N, Greeff, J, Hebart, M, Hiendleder, S, Edwards, JH, Kelly, J, Kind, K, Kleemann, D (2014) *Anim. Prod. Sci.* **54:** 667.

Brown, D, Jones, R, Hinch, G (2014). *Anim. Prod. Sci.* **54**: 736.

Butler, D, Cullis, BR, Gilmour, A, Gogel, B (2009) *The State of Queensland, Department of Primary Industries and Fisheries.*

Hatcher, S., K. D. Atkins, and E. Safari. (2014) *J. Anim. Sci.* **87:** 2781.

MLA (2012) Australian Goat Industry RD&E Strategy – 2012. *Meat and Livestock Australia*.

Singh, D., Mishra, H., Singh, C., (1990) *Anim. Prod*. **51:** 559.

Snyman, M. (2010) *S Afr J Anim Sci* **40**: 54.

Walkden-Brown, S., Bocquier, F., (2000) *Proc. 7th International Conference on Goats* 389.

Zhang, C., Yang, L., Shen, Z. (2008) *Livest. Sci.* **115:** 73.

Zhang, C.-Y., Chen, S.-L., Li, X., Xu, D.-Q., Zhang, Y., Yang, L.-G., (2009) *Livest. Sci.* **125:** 60.

GENETIC RELATIONSHIPS BETWEEN BREECH COVER, WRINKLE AND LAMB SURVIVAL IN MERINO SHEEP

S. Hatcher[1,2], D.J. Brown[1,3], F.D. Brien[1,4] and M.L. Hebart[1,4]

[1]CRC for Sheep Industry Innovation, Homestead Building, UNE, Armidale, NSW, 2351
[2]NSW DPI, Orange Agricultural Institute, Forest Road, Orange, NSW, 2800
[3]Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW, 2351
[4]University of Adelaide, School of Animal and Veterinary Sciences, Roseworthy, SA, 5371

SUMMARY

Phenotypic and genetic correlations (r_g) between breech cover and wrinkle scores with lamb survival were estimated from research (Cooperative Research Centre for Sheep Industry Innovation (Sheep CRC) Information Nucleus (IN)) and industry data (Sheep Genetics (SG) MERINOSELECT). Breech cover had low to moderate antagonistic r_g (0.24 to 0.53) with direct lamb survival, despite favourable r_g with some lamb survival indicator traits. While there was some inconsistency in the direction of r_g at the various age expressions of wrinkle and direct lamb survival, the r_g between wrinkle and lamb survival tended to be negative and favourable.

INTRODUCTION

High wrinkle scores are related to poorer reproductive outcomes (Turner and Young 1969). Folds$^+$ ewes, selected on total fold score at four months of age, weaned half as many lambs during their lifetime as Folds$^-$ ewes (Dun 1964). Folds$^+$ rams were less fertile and sired fewer lambs than Folds$^-$ rams irrespective of whether they were mated to Folds$^+$ or Folds$^-$ ewes (Dun and Hamilton 1965). In addition, more Folds$^+$ single-born lambs died compared with Folds$^-$ singletons due to dystocia resulting from increased birthweight together with the smaller size of the Folds$^+$ ewe (Dun and Hamilton 1965). While these early studies provided insight into the relationship between wrinkle and reproduction, few studies since have reported phenotypic and genetic correlations between these traits and none have included breech cover. The recent focus by some Merino breeders to genetically reduce wrinkle, to decrease flystrike and the need for mulesing (James 2006), is likely to generate correlated improvements in Merino reproductive performance. This paper reports estimates of phenotypic and genetic correlations between wrinkle and breech cover with lamb survival to weaning (a key component of reproductive performance), assessed as a trait of both the ewe and the lamb, using a combination of industry and research data.

MATERIALS AND METHODS

Research data. Data from Merino progeny born into the Sheep CRC's IN (van der Werf *et al.* 2010) between 2007 and 2011 were used. Twice daily lambing rounds were conducted from the start of lambing in each year to uniquely identify each lamb, designate a dam, confirm birth status (alive or dead, LS0), collect early life information (sex and type of birth) and record birth weight (BWT), birth coat score (BCS), overall birth vigour (OBV), lambing ease (LE), rectal temperature (RT), three skeletal measures (crown rump length CRL, metacarpal length ML and thorax circumference THO) and a suite of timed lamb behaviours (time taken from lamb release after tagging to bleat BLT, stand STD, contact the ewe CONT, contact the udder UDD and follow the ewe FOLL) as described by Brien *et al.* (2015). Identities of all surviving lambs were confirmed at 3 (LS3) and 7 days of age (LS7) and at weaning (approx. 91 days of age, LSURV). Breech cover

* AGBU is a joint institute of NSW Department of Primary Industries and The University of New England

(BCOV), breech (BRWR), body (BDWR) and neck (NKWR) wrinkle were scored according to industry standards (AWI Ltd and MLA Ltd 2013) at marking (7-43 days, BCOV and BRWR only) as yearlings (11 months) and adults (23 months). Variance components were estimated using ASReml (Gilmour *et al.* 2009) from a series of univariate animal models (see Brien *et al.* (2015) for the lamb survival traits and Hatcher and Preston (2015) for BCOV and wrinkle traits). Phenotypic and genetic correlations between the lamb survival, BCOV and wrinkle traits were estimated from appropriate covariances using a series of bivariate analyses with the residual covariance with the direct lamb survival traits set to 0.

Industry data. Pedigree and performance data were extracted from the SG MERINOSELECT database (Brown *et al.* 2007). A subset of 29 flocks was selected based on their recording of wrinkle and reproduction traits and were a mix of industry ram breeders, research and sire evaluation flocks. From within these flocks all animals with both sire and dam pedigree, and born since 1990 were included. Data were extracted for all early breech wrinkle (EBWR, marking to weaning), late body wrinkle (LBDWR, yearling to adult), lamb survival as a trait of the lamb (LSURV) and lamb survival as a trait of the ewe (SURV) (Table 2). The pedigree was built using all available ancestral information. For the bivariate analyses there was between 270 and 607 sires with records for both traits depending on the trait combination. Parameters were estimated in bivariate animal model analyses in ASReml (Gilmour *et al.* 2009). For EBWR and LBDWR fixed effects of contemporary group (defined as flock, year of birth, sex, date of measurement, management group subclass), birth type, rearing type, age of dam, and animal's age at measurement were fitted. For LSURV and SURV contemporary group (defined as flock, year of lambing, ewe age class (1, 2 and 3+) and management group) was the only fixed effect fitted. Random terms for direct genetic effects and sire by flock-year were fitted for all traits. Maternal permanent environment was included for wrinkle and LSURV and a repeated record term added for SURV. For correlations involving LSURV, the residual covariance was fixed to zero as only surviving lambs had records for both traits. Genetic groups were fitted in all analyses and allocated on a flock basis for linked flocks with sufficient data (Swan *et al*. 2015).

RESULTS AND DISCUSSION

Research data. Some of the r_g, particularly those involving direct lamb survival, had large standard errors (Table 1). This may be due to the dataset size, the number of sire families represented or the low additive variation for the direct lamb survival traits (h^2<0.01, Brien *et al.* 2015). Nevertheless, until such time as lamb survival records become more widely represented in industry data or further analysis is done on pooled resource flock data, these results are likely to remain the only available estimates.

There were unfavourable r_g between marking, yearling and adult BCOV and direct lamb survival (Table 1), such that selection for increased bare area around the perineum and breech will generate correlated decreases in lamb survival. The low negative r_g between BWT and marking BCOV implies that selection for bare breeches is genetically associated with higher BWT, while the low positive r_g of yearling and adult BCOV with RT may be indicative of decreased thermoregulation capacity. There were favourable r_g between BCOV and LE as well as some timed lamb behaviours, no significant r_g between BCOV and any of the skeletal measurements and the direction of the r_g with BCS, OBV and the timed lamb behaviours varied with age. However, as RT had the strongest r_g with LSURV among the suite of indicator traits (Brien *et al.* 2015), it is likely to dominate the genetic relationships with lamb survival despite the favourable relationships between BCOV and the other indicator traits.

The medium to high negative rg between the marking and yearling expressions of BRWR with direct lamb survival were indicative of a favourable relationship. So selection for plainer breeches will generate correlated improvements in lamb survival. Plainer breeches were also genetically

associated with lower BWT, smaller skeletal measurements and higher RT. However, the r_g between the age expressions of BRWR with OBV and the timed lamb behaviours varied in direction. The r_g between the various lamb survival traits and yearling BRWR and BDWR were similar as were those between adult NKWR and BRWR. Due to the nature of the IN data set, our ability to fully account for across flock differences via fitting genetic groups was limited. As a result, the r_g may be inflated compared to what would be observed within other flocks.

Table 1. Genetic correlations between a) marking and yearling and b) adult of breech cover (BCOV), breech (BRWR), body (BDWR) and neck (NKWR) wrinkle with lamb survival and various indirect selection criteria for lamb survival in the Sheep CRC IN

High r_g (≥0.6) are highlighted in **bold** and medium r_g are underlined. The r_g for LS0, BWT, CRL and BLT have been omitted from 1b) as they were all negligible (|r|<0.2).

a)	Marking		Yearling			
	BCOV	BRWR	BCOV	BRWR	BDWR	NKWR
LS0	-0.30±0.08	-0.03±0.05	0.53±0.13	-0.20±0.06	-0.37±0.06	0.16±0.07
LS3	0.24±0.22	-0.42±0.14	0.34±0.30	-0.41±0.17	-0.59±0.16	-0.06±0.18
LS7	0.49±0.25	-0.48±0.16	0.43±0.31	-0.49±0.18	-0.59±0.16	-0.16±0.18
LSURV	0.49±0.68	**-0.99±0.71**	0.21±0.63	**-0.98±0.87**	**-0.82±0.63**	-0.39±0.47
BWT	-0.27±0.12	0.38±0.08	-0.06±0.19	0.12±0.10	0.29±0.10	-0.01±0.11
BCS	-0.37±0.13	**-0.60±0.08**	0.32±0.21	-0.08±0.11	-0.11±0.11	-0.11±0.12
OBV	0.38±0.16	-0.30±0.11	-0.38±0.27	0.10±0.14	0.06±0.14	-0.28±0.15
LE	0.26±0.26	-0.38±0.19	-0.08±0.38	-0.17±0.22	-0.32±0.22	-0.48±0.23
RT	0.13±0.26	-0.36±0.18	0.35±0.36	-0.08±0.22	0.14±0.22	0.17±0.23
CRL	0.04±0.13	0.20±0.08	-0.04±0.19	0.11±0.10	0.18±0.09	-0.02±0.11
ML	0.12±0.19	0.32±0.13	0.16±0.26	0.33±0.14	0.39±0.13	0.27±0.15
THO	-0.13±0.16	0.31±0.09	0.10±0.22	0.22±0.11	0.38±0.11	0.32±0.13
BLT	-0.05±0.21	-0.21±0.15	0.57±0.30	-0.08±0.17	0.01±0.17	0.10±0.19
STD	0.01±0.31	-0.36±0.23	**0.73±0.37**	0.09±0.23	-0.10±0.22	0.05±0.25
CONT	-0.25±0.22	-0.22±0.14	-0.08±0.28	-0.11±0.15	-0.15±0.15	-0.19±0.17
UDD	-0.03±0.51	-0.00±0.39	0.46±0.61	-0.11±0.41	-0.37±0.44	**-0.69±0.48**
FOLL	-0.52±0.28	-0.37±0.22	0.23±0.39	-0.09±0.23	-0.22±0.24	-0.08±0.26

b)	Adult			
	BCOV	BRWR	BDWR	NKWR
LS3	0.17±0.36	0.04±0.19	0.06±0.23	0.44±0.28
LS7	0.34±0.37	0.03±0.20	-0.01±0.23	0.44±0.28
LSURV	-0.14±0.72	0.10±0.39	-0.07±0.46	**0.64±0.66**
BCS	0.44±0.23	-0.09±0.13	0.00±0.15	0.06±0.18
OBV	0.39±0.25	0.15±0.15	-0.38±0.28	-0.05±0.22
LE	**0.71±0.46**	-0.03±0.04	-0.30±0.28	**-0.71±0.32**
RT	0.22±0.40	-0.09±0.23	0.03±0.29	0.12±0.33
ML	-0.06±0.27	0.10±0.17	0.02±0.19	0.01±0.22
THO	0.01±0.22	0.46±0.14	0.29±0.16	0.57±0.18
STD	**0.66±0.35**	0.32±0.32	0.11±0.30	-0.03±0.32
CONT	0.07±0.27	0.23±0.18	0.10±0.20	0.06±0.23
UDD	**0.64±0.73**	-0.38±0.49	0.13±0.48	-0.24±0.53
FOLL	0.58±0.34	0.09±0.27	0.05±0.28	-0.30±0.31

Industry data. EBWR and LBDWR were moderately heritable and had small but significant maternal effects (Table 2) consistent with IN estimates (Hatcher and Preston 2015). The heritability of LSURV was low (0.02) with a significant maternal effect (0.14) which agreed with IN estimates (Brien *et al.* 2015). The r_g between LSURV and wrinkle were small and negative but

not significantly different from zero (Table 2). The IN results were generally stronger negative, but also often not significantly different to zero. However, SURV was more negatively correlated with wrinkle both at the genetic and maternal permanent environment levels than LSURV.

Table 2. Phenotypic variance (σ^2_p), maternal permanent environment (PE), direct heritability (diagonal bold type), genetic correlations (below diagonal), maternal permanent environmental correlation (above diagonal) between wrinkle and lamb survival traits

	EBWR	LBDWR	LSURV	SURV
Records	80,467	38,348	46,826	11,957
σ^2_p	0.59±0.00	0.51±0.00	0.11±0.00	0.09±0.00
PE	0.04±0.01	0.00±0.01	0.14±0.01	NE
EBWR	**0.32±0.01**	NE	-0.14±0.06	NE
LBDWR	NE	**0.40±0.01**	0.99±0.79	NE
LSURV	-0.05±0.10	-0.24±0.20	**0.02±0.01**	NE
SURV	-0.17±0.11	-0.28±0.18	NE	**0.03±0.01**

NE: not estimated or not estimable

CONCLUSIONS

Merino breeding programs with an emphasis on reducing wrinkle are likely to generate correlated improvements in reproduction both as a trait of the ewe and the lamb. While breeders using BCOV as an indirect selection criterion for flystrike could expect correlated decreases in direct lamb survival despite favourable genetic relationships with some lamb survival indicator traits. However the inconsistency between age expressions of the wrinkle and BCOV traits due to low precision of the genetic correlations does add some uncertainty to these conclusions. The inaccurate parameter estimates are due to some animals not having both records, as dead lambs were not scored for BCOV or wrinkle and we therefore rely on the genetic relationships between animals via the pedigree to estimate these genetic correlations. Accurate parameters and multiple trait index predictions are required to properly quantify the impact of these relationships for Merino breeding programs.

REFERENCES

AWI Ltd and MLA Ltd (2013) *Visual Sheep Scores - Researcher Version.* Sydney, Australia: AWI Ltd & MLA Ltd.

Brien F.D., Hebart M.L, Hinch, G.N., Brown, D.J. and van der Werf J.H.J. (2015) *Anim. Prod. Sci.* Submitted.

Brown D.J., Huisman A.E., Swan A.A., Graser H.U., Woolaston R.R., Ball A.J., Atkins K.D. and Banks R.G. (2007) *Assoc. Advmt Anim. Breed. Genet.* **17:** 187.

Dun R.B. (1964) *Aust. J. Exp. Agric. Anim. Husb.* **4:** 376.

Dun R.B. and Hamilton B.A. (1965) *Aust. J. Exp. Agric.* **5:** 236.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) *ASReml User Guide Release 3.0.* Hemel Hempstead, UK: VSN International Ltd.

Hatcher S. and Preston J.W.V. (2015) *Small Rumin. Res.* Submitted.

James P.J. (2006) *Aust. J. Exp. Agric.* **46:** 1.

Swan A.A., Brown D.J. and van der Werf J.H.J. (2015) *Anim. Prod. Sci.* dx.doi.org/10.1071/AN14560.

Turner H.N. and Young S.S.Y. (1969) *Quantitative genetics in sheep breeding*. North Melbourne, Australia:Macmillan of Australia.

van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Anim. Prod. Sci.* **50:** 998.

A GENOME-WIDE ASSOCIATION STUDY OF NON-ADDITIVE EFFECTS FOR MILK YIELD AND FERTILITY IN HOLSTEIN AND JERSEY DAIRY CATTLE

H. Aliloo[1,2,3], J.E. Pryce[1,2,3], O. González-Recio[1,3], B.G. Cocks[1,2,3] and B.J. Hayes[1,2,3]

[1] Biosciences Research Division, Department of Economic Development, Jobs, Transport and Resources, AgriBio, 5 Ring Road, Bundoora, VIC. 3083, Australia
[2] School of Applied Systems Biology, La Trobe University, Bundoora, VIC. 3083, Australia
[3] Dairy Futures Cooperative Research Centre (CRC), AgriBio, Bundoora, VIC. 3083, Australia

SUMMARY

It has been suggested that traits associated with fitness, such as fertility, may have proportionately more genetic variation arising from non-additive effects than traits with higher heritability, such as milk yield. Here, we performed a large genome scan with 408,255 single nucleotide polymorphism (SNP) markers to identify chromosomal regions associated with dominance and epistatic (pairwise additive × additive) variability in milk yield and fertility (measured by calving interval), using 7,055 genotyped and phenotyped Holstein cows. The results were subsequently replicated in an independent set of 3,795 Jersey cows. We identified genome regions with replicated dominance effects for milk yield on *Bos taurus* autosomes (BTA) 2, 3, 5, 26 and 27 whereas SNPs with replicated dominance effects for fertility were found on BTA 1, 2, 3, 7, 23, 25 and 28. A number of significant epistatic effects for milk yield on BTA 14 were found across breeds. However, these were likely to be associated with the mutation in the *diacylglycerol O-acyltransferase 1* (*DGAT1*) gene, given that the associations were no longer significant when the full additive effect of the *DGAT1* mutation was included in the epistatic model. The results of our study suggest that individual non-additive effects make a small contribution to the genetic variation of milk yield and fertility.

INTRODUCTION

Female fertility is of great interest to the dairy industry because impaired reproductive ability can reduce the profitability of a dairy herd, through increased expense of additional inseminations, veterinary treatments and replacement cows. Selection to improve milk production traits in Holstein and Jersey cattle populations has led to a decline in fertility traits in the last few decades due to unfavourable genetic correlations between fertility and milk production (Berry *et al*. 2014). Many countries have now included fertility in their national breeding goals. However fertility related traits usually have low heritability estimates and genetic improvement through traditional breeding programs is slow, although substantial genetic variation exists (Khatkar *et al*. 2014). For traits such as fitness traits, where heritability estimates are low the non-additive part of genetic variation could be used to genetically improve the trait of interest. Non-additive genetic variation is the result of allele by allele interactions and involves intra-locus (dominance) and inter-locus (epistasis) interactions.

An increasing availability of genotypes coupled with phenotypes has provided a new opportunity for estimation of non-additive genetic effects. Genome-wide association studies can be used to estimate both the additive and non-additive effects of genetic markers, but most published GWAS for dairy cattle to date have focused on additive effects of genes while non-additive interactions are generally neglected.

The objective of this study was to detect chromosomal regions with non-additive effects for calving interval (CI) and milk yield (MY) using a Holstein cow discovery population. We then attempted to validate these associations in an independent Jersey population of cows.

MATERIALS AND METHODS

Data. Animals were genotyped with Illumina BovineSNP50 v2 BeadChip (Illumina, San Diego, CA, USA) and their 50K SNP data were imputed to the high density (HD) 800k panel using Beagle 3 (Browning and Browning 2009). Standard quality control checks were applied on genotypic data prior to the imputation step. Accurate estimation of dominance effects of the SNPs requires enough observations in all three classes of SNP genotypes. Therefore, a further 223,748 SNPs were removed from HD SNP panel due to a genotype class with frequency < 0.01 in both Holstein or Jersey animals. The final set comprised 408,255 SNPs.

Phenotypic data included 23,198 and 11,091 milk yield and calving interval records respectively for 7,055 Holstein and 3,795 Jersey cows (some cows had records across multiple lactations). These records were pre-corrected for herd-year-season, age at calving, parity and month of calving using a fixed model on full national data set of phenotypes. Residuals from this model were then used as the response variable in GWAS analyses for the genotyped animals.

Statistical model. The mixed linear model used was:

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wpe} + \mathbf{e}$$

where **y** is a vector of phenotypes (CI or MY), $\mathbf{1_n}$ is a incidence vector of ones, μ is the population mean term, **b** is the vector containing relevant additive or dominance marker effects as specified below, **u** contains polygenic effects assumed to be distributed as $\mathbf{u} \sim \mathrm{N}(0, \mathbf{A}\sigma_g^2)$ with **A** being the pedigree based numerator relationship matrix, **pe** is the vector of random permanent environmental effects with $\mathbf{pe} \sim \mathrm{N}(0, \mathbf{I}\sigma_{pe}^2)$ and **e** is a vector of random residual deviates distributed as $\mathbf{e} \sim \mathrm{N}(0, \mathbf{I}\sigma_e^2)$. Then **X** is a design matrix allocating records to marker effects (dominance or additive by additive) and **Z** and **W** are incidence matrices for the random effects. σ^2_g, σ^2_{pe} and σ^2_e are polygenic additive, permanent environmental and residual variances, respectively.

The original marker covariates (0, 1 or 2) were corrected for allele frequencies (Vitezica *et al.* 2013) to build **X**, so that $x_{ij(a)}$ = {$-2p$, $(q-p)$ or $2q$} for additive effects of aa, Aa and AA genotypes, respectively, with p and q being the frequencies of alleles A and a at marker j in the population. For dominance effects, aa, Aa and AA genotypes were coded as $x_{ij(d)}$ ={$-2p^2$, $2pq$ and $-2q^2$}. Then, the contents of **Xb** varied with the type of the genetic effect being tested. For dominance, $\mathbf{Xb} = \{x_{ij(a)}a_j + x_{ij(d)}d_j\}$, where d_j is the dominance effect of marker j. In the epistasis model, $\mathbf{Xb} = \{x_{ij(a)}a_j + x_{ik(a)}a_k + x_{ijk(e)}a_{jk}\}$, where $x_{ijk(e)}$is the qualification for nested interaction effects involving markers j and k, a_k is the additive effect for the k marker and a_{jk} is the pairwise additive by additive epistatic marker effect between markers j and k.

Validation. To confirm if significant SNPs were consistent between breeds, results from the larger Holstein population (discovery set) were validated in the Jersey breed in two different ways. First, if a significant SNP was found in the discovery process, we examined whether it was also significant in the validation population. Second, for each significant SNP in the discovery population, we searched for significant SNPs in the validation population within the region 500 kb downstream and upstream of the identified SNP.

The false discovery rate was calculated following the approach proposed by Bolormaa *et al.* 2010 as: $\%FDR = (P(1 - S/T)/((S/T)(1 - P))) \times 100$
where P is the P-value threshold in F-test, S is the number of significant SNPs according to this threshold and T is the total number of tests.

For dominance models, all of the markers in the final HD panel were used. To reduce the dimension of SNP combinations tested in the epistatic models, only significant SNPs determined using the P-value of the F-test of their additive effects in the Holstein discovery set were included.

RESULTS AND DISCUSSION

Dominance. The false discovery rate (FDR) for dominance effects were high, at 100% for both traits (Table 1) meaning that the number of significant SNPs in Holsteins is smaller than expected by chance. Forty SNPs were significant ($P< 0.0001$) in the Holstein discovery population for MY. Only 1 of these was also significant ($P < 0.01$) in Jersey cows, but with different signs observed in the discovery and validation analyses and with a FDR of 39 % (Table 1). For CI, 36 SNPs were found to have significant ($P < 0.0001$) dominance associations in the Holstein discovery set (Table 1). Of these, 3 (1 with same direction) SNPs were validated in individual validation (FDR = 11 %).

The segment validation approach was more successful; the number of validated SNPs for MY increased to 21 ($P < 0.01$) with a FDR in the validation population being equal to 1%; 10 SNPs were validated for CI (FDR = 3 %) within discrete regions. The validated SNPs with significant dominance effects on MY and CI were detected on 5 (BTA 2, 3, 5, 26 and 27) and 7 (BTA 1, 2, 3, 7, 23, 25 and 28) chromosomes, respectively (Figure 1).

Table 1. *P*-value thresholds and the number of SNPs with significant dominance effects and corresponding false discovery rates (FDR) for milk yield (MY) and calving interval (CI)

	Discovery			Individual validation				Segment validation		
Trait	*P*	Holstein	FDR (%)	*P*	Jersey	FDR (%)	Same Dir.	*P*	Jersey	FDR (%)
MY	0.0001	40	102	0.01	1	39	0	0.01	21	1
CI	0.0001	36	113	0.01	3	11	1	0.01	10	3

[1]Number of same direction SNP effects in discovery and validation populations

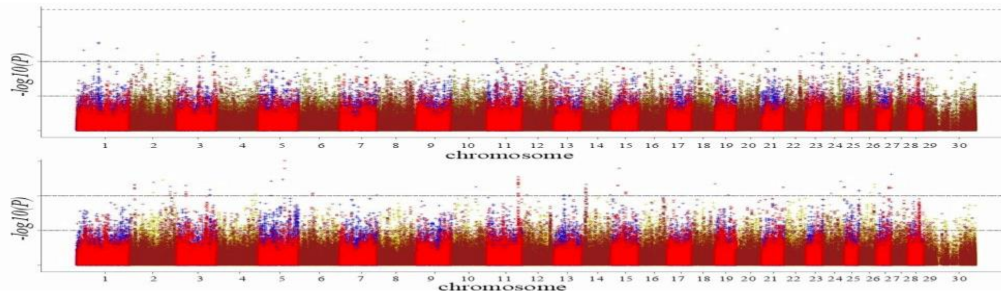


Figure 1. Manhattan plot of dominance SNP effects for fertility (top) and milk yield (bottom) with chromosome number on horizontal axis and –log10(*P*-value) on vertical axis.

Epistatic interactions. A larger number of significant pairwise interactions were found for milk yield compared to fertility (Table 2). For MY there were 3,700 significant pairwise interactions in the discovery population of Holstein cows at the threshold of $P < 10^{-7}$. Of which 165 were validated ($P < 1\times10^{-5}$) in the Jersey population (Table 2). The number of validated additive × additive effects that were in the same direction in both Holstein and Jersey data was 163 out of 165. In all epistasis analyses of MY, FDRs were calculated to be very close to zero. Since all of the SNPs that had validated interactions for MY were located at the beginning of BTA 14 and near the *DGAT1* gene, we suspected that these interactions may have been due to the *DGAT1* mutation effect. Therefore, the epistatic model was extended to include a SNP in the *DGAT1* gene itself as a fixed effect to see if any of the interactions remained significant. The absence of significant interactions in this region after including the *DGAT1* effect in the model suggests that the identified significant pairwise interactions identified were picking up the *DGAT1* effect by creating haplotype like combinations. That is, the linkage disequilibrium of SNP allele

combinations with the *DGAT1* mutation was higher than for the individual SNP, rather than a true epistatic interaction.

Five additive × additive interactions were found significant ($P < 0.0001$) for CI in Holsteins with a FDR of 18%. However, none of these was validated ($P < 0.01$) in the Jersey population.

Table 2. *P*-value thresholds, number of significant pairwise additive × additive interactions and calculated false discovery rates (FDR) for milk yield (MY) and calving interval (CI)

		Discovery			Validation			
Trait	No. of interactions	*P*	Holstein	FDR (%)	*P*	Jersey	FDR (%)	Same Dir.[1]
MY	255,255	10^{-7}	3700	0	10^{-5}	165	0	163
CI	9,180	10^{-4}	5	18	0.01	0	NA	NA

[1]Number of same direction SNP effects in discovery and validation populations

Implications. One critical parameter determining the power of a GWAS is the amount of LD between the observed SNP and the unobserved causal variant. The success of a GWAS in identifying QTLs with additive effects is controlled by r^2 (r is the correlation between genetic marker and causative mutation) while detection of dominance or pairwise additive by additive effects depends on r^4. This indicates a much higher reliance on LD in searching for non-additive effects compared to additive effects, if LD between the markers and QTL is incomplete (Wei *et al.* 2014). This, and possibly the relatively small size of individual dominance and epistatic effects, was reflected in results of this study in which a larger number of additive markers were identified than the markers with dominance and epistasis effects for both traits under investigation.

The standard in reporting GWAS results is validation and before genotype-phenotype relationships can be used in selection decisions, they should be replicated in an independent population to confirm generalized effects in multiple populations. Validation of GWAS results across breeds can refine QTL regions to narrower intervals and is powerful in identifying positional candidate genes. This is because the extent of LD across cattle breeds is limited in contrast to within a breed, where considerable LD can be maintained in intervals up to 1 Mbp as a result of a relatively small effective population size. We validated a lower number of non-additive genetic associations than additive effects such that very few dominance effects for MY and CI were confirmed and no epistasis effect was common across Holstein and Jersey cows for CI. This trend is in agreement with the fact that the higher dependence on LD in searching for dominance and epistatic effects compared to additive effects significantly decreases the chance of validating associations in independent populations for non-additive effects of the markers (Wei *et al.* 2014).

CONCLUSION

We identified and validated a small number of SNPs with suggested dominance effects on MY and CI in Australian Holstein and Jersey cows. Given our results, identifying non-additive gene actions using single SNP regression in a GWAS setting will require very large datasets to capture the likely very small individual non-additive genetic effects.

REFERENCES

Berry D.P., Wall E., Pryce J.E. (2014) *Animal.* 8:105.
Bolormaa S., Pryce J.E., Hayes B.J., Goddard M.E. (2010) *J. Dairy Sci.* **93**: 3818.
Browning B.L., Browning S.R. (2009) *Am. J. Hum. Genet.* **84**: 210.
Khatkar M.S., Randhawa I.A.S., Raadsma H.W. (2014) *Livest. Sci.* **166**: 144.
Vitezica Z.G., Varona L., Legarra A. (2013) *Genetics.* **195**: 1223.
Wei W-H., Hemani G., Haley C.S. (2014) *Nat. Rev. Genet.* **15**: 722.

EFFECTS OF SELECTION FOR FERTILITY ON LACTATION CURVES

E.M. Strucken[1], Y.C.S.M. Laurenson[1] and G.A. Brockmann[2]

[1] School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia
[2] Breeding Biology and Molecular Genetics, Humboldt-Universität zu Berlin, Berlin, Germany

SUMMARY

Breeding indices have enabled farmers to select for multiple traits simultaneously, including negatively correlated traits such as milk production and fertility. This negative correlation is believed to be either caused by an energy deficit during early lactation or serves a functional purpose in providing optimal birth spacing. A linear regression was carried out between parameters describing a lactation curve and a fertility index (RZR) and milk yield EBVs (mEBVs) to determine the effects of selection on the lactation curve. Breeding values of first lactation milk yield and a RZR were available for 2,405 sires. Additionally, these sires had test-day records of the first lactation of ~2M daughters. There was a negative correlation between mEBVs and RZR ($r=-0.27$, $P<0.0001$). Selection for fertility resulted in higher initial milk yield with an early peak yield. This suggests that an early peak occurs to provide offspring with sufficient milk despite a potential energy deficit. Further, an early peak provides an increased duration over which milk production declines and therefore sufficient time for the cow to recover from the energy deficit prior to a subsequent pregnancy. Finally, current production environments could be optimised to fulfil the genetic potential of high producing dairy cows.

INTRODUCTION

Negative genetic correlations have been reported between milk production and a variety of functional traits (Dekkers *et al.* 1998; Muir *et al.* 2004) including health and fertility (Ingvartsen *et al.* 2003; Oltenacu and Broom 2010). Both causal and functional hypotheses have been proposed to account for such negative correlations (Strucken *et al.* 2015). Collard (2000) proposed that the energy deficit experienced during early and peak lactation causes detrimental effects on health and fertility. However, very low correlations have been reported between total milk yield and energy balance (Spurlock *et al.* 2012). Alternatively, the negative impact of lactation on fertility may serve a functional purpose to provide optimal birth spacing for the survival of offspring.

Whilst milk production predominantly remains the most economically important trait for dairy farmers, functional traits such as conformation, udder health and fertility have become more prominent as the importance of animal welfare and longevity has increased (VanRaden, 2004; Miglior *et al.* 2005). Consequently, breeding goals were adjusted to incorporate health and fertility traits into breeding indices (Osteras *et al.* 2007; Boichard and Brochard 2012). These breeding indices allow traits to be weighted according to their economic importance and heritability, and account for phenotypic and genetic correlations between traits (Dekkers 2007). As such, this has enabled dairy farmers to breed for milk production and functional traits without requiring knowledge on how these practices impact upon the dynamics of milk production.

The dynamics of milk production can be described using an appropriate lactation model such as the Wilmink curve (Wilmink 1987). By fitting such a model to milk yield test-day records, a lactation cycle can be summarised using a minimal number of parameter values. These parameter values can subsequently be used to perform a linear regression with estimated breeding values for total milk yield and a fertility index. Understanding the impact of selection for fertility upon the parameter values of a lactation model could aid in determining whether the negative correlation between production and fertility is causal or functional. If the observed impact of production upon

fertility is caused by the energy deficit experienced during early and peak lactation, then a reduced peak milk yield would be expected to occur at a later time point. However, if the negative correlation serves a functional purpose to provide optimal birth spacing for the survival of offspring, then a reduced persistency would be expected.

MATERIALS AND METHODS

Data. Estimated breeding values (EBVs) for 2,405 Holstein Friesian sires, and test-day records for 1,797,852 daughters, were provided by VIT, Verden (Germany). Each bull had an average of 747 daughters (min=50, max=84,387).

Fertility breeding values were pre-corrected for *herd*year*, *month of insemination*, *age at insemination*, *parity*age* at insemination, *status* and *effect of the bull* (c.f. VIT April 2015). Breeding values were available for six measures: *non-return rate* 56 days post-insemination (separated for heifers and cows), *first to successful insemination* (separated for heifers and cows), *calving to first insemination*, and *days open*. The fertility EBVs were summarized in a fertility index (RZR) which was standardized to a mean of 100 and a standard deviation of 12. Further, corrected 305d EBVs for milk yield (mEBV) were available for the first lactation. These breeding values were raw values, representing actual yield deviations from the population mean.

Daughter records comprised the first lactation with an average of 8 test-day records per cow (min=1, max=20). The average lactation length was 259 days (min=5, max=330).

Table 1. Data description for the fertility index (RZR) and 305d milk yield EBVs (mEBV) of the first lactation for sires, and test-day milk yield of daughters

	N	Mean	Min	Max	SD
RZR	2,405	100.95	62.00	136.00	9.90
RZR (top 10 sires)	10	130.4	128.00	136.00	2.37
RZR (worst 10 sires)	10	71.6	62.00	75.00	4.01
mEBV (kg)	2,405	711.03	-1438	2774	609.43
mEBV (top 10 sires)	10	2583	2408	2774	118.51
mEBV (worst 10 sires)	10	-1118.6	-1438	-946	157.77
milk yield total (kg)	14,862,232	25.57	2.00	98.80	6.54
milk yield per sire (kg)	386-731,431	18.9-31.4	2.0-12	30.6-98.8	4.4-8.8

Analyses. Pearson's correlation coefficients were calculated between RZR and mEBVs. Wilmink curve parameters were estimated per sire with a non-linear model in *R*. The Wilmink curve is given as (Wilmink 1987):

$$y = a + b * exp^{-k*DIM} + c * DIM$$

Y is the test-day record; *a* is the potential maximum daily milk yield (kg); *b* determines the y-intercept (y-intercept = *a*+*b*); *c* is the gradient of the linear decay in milk yield (kg d^{-1}); k is the growth rate. Parameters *b* and *c* are both negative for a lactation curve. Convergence was achieved for 2,392 sires. Each parameter was subsequently used for a linear regression with RZR and mEBVs.

Environmental factors such as season or age are known to impact upon milk production. To determine the effect of environmental factors, we estimated the Wilmink curve parameters in a linear mixed model including the fixed effects of *age at calving*, *year season*, and *milk recording system* nested within *farm*. Sire was included as a random effect. These calculations were carried out across the top 10 and worst 10 sires for RZR and mEBVs, to provide the greatest contrast

between the production curves of high and low ranked sires (**Table 1**). In order to fit a linear model, Wilmink parameter *k* was fixed at 0.06316 (top 10 RZR), 0.05912 (worst 10 RZR), 0.05258 (top 10 mEBVs), and 0.07269 (worst 10 mEBVs) based on preliminary calculations.

RESULTS

The pseudo-genetic correlation between mEBVs and RZR was significantly negative (r=-0.27, P<0.0001), confirming previous reports (Oltenacu and Broom 2010). As expected, this negative correlation resulted in a negative association between Wilmink curve parameter *a* (potential maximum) and RZR, and a positive association with mEBVs (**Table 2**). Further, increases in both RZR and mEBV resulted in a significant reduction in parameter *b*, and hence a higher y-intercept.

Table 2. Effects of fertility index (RZR) and mEBVs on lactation curve parameters

	RZR ±se	R^2 (adj)	mEBV ±se	R^2 (adj)
a	-0.042 ±0.005***	0.03	0.0023 ±0.00***	0.36
b	0.031 ±0.008***	0.005	0.0009 ±0.00***	0.02
c	0.000006 ±0.00	0.0003	-0.0000003 ±0.00	0.0008
k	0.0002 ±0.00**	0.003	-0.00001 ±0.00***	0.06

*** P<0.0001 **P<0.001

a: potential maximum daily milk yield; *b*: determines y-intercept; *c*: gradient of the linear decay in milk yield; *k*: growth rate

Effects of RZR and mEBVs on curve parameter *c* (determining the gradient of the linear decay in milk yield) were not significant. However, a reduction in parameter *c* is indicated for increasing RZR (i.e. decreased decay rate), whilst increasing mEBV caused an increase in parameter *c* (i.e. increased decay rate, **Table 2**). Parameter *k* (growth rate) increased for better RZR, and decreased for better mEBVs (**Table 2**).

Correction for environmental effects had an impact upon lactation curve parameters (**Table 3**). The largest impact was observed for parameter *a* where correction for environmental effects increased the potential maximum daily milk yield for the best mEBV sires, and decreased for the worst RZR and mEBVs. Further, correction for environmental effects reduced parameter *b* for the best mEBVs, causing a higher y-intercept (**Table 3**).

Table 3. Corrected and uncorrected Wilmink parameters based on best and worst sires

	RZR		mEBVs	
	10 best	10 worst	10 best	10 worst
	corrected parameters			
a	30.93	30.82	54.14	23.77
b	-10.74	-13.70	-1.3E-07	-9.66
c	-0.038	-0.036	-0.020	-0.036
k	0.063	0.059	0.053	0.073
	uncorrected parameters			
a	30.82	33.07	34.81	28.80
b	-10.80	-13.85	-13.70	-9.61
c	-0.038	-0.037	-0.036	-0.036
k	0.063	0.059	0.053	0.073

DISCUSSION

The functional and causal hypotheses previously proposed to account for the negative correlation between production and fertility may be expected to impact upon the shape of the lactation curve in different manners. Reductions in fertility caused by an energy deficit experienced during early and peak lactation would be expected to impact upon early lactation. Whereas, if the observed reduction in fertility serves a functional purpose to provide optimal birth spacing, then an increased rate of decay (decreased parameter *c*) would be expected.

All significant impacts of the fertility index occurred for parameters determining early and peak lactation (*a*, *b*, and *k*), whilst no significant effect was found on rate of decay (parameter *c*). This is consistent with previous studies which reported genetic loci affecting early and peak

lactation (Strucken *et al.* 2011; Strucken *et al.* 2012). These results appear to support a causal hypothesis where the negative correlation between production and fertility is due to an energy deficit in early lactation. The expectation was that the fertility index would cause a decreased potential maximum milk yield and a slower growth rate. However, the results showed an increased initial milk yield (y-intercept) and an increased growth rate causing an early production peak. The increased initial milk yield and growth rate during early lactation occurs to provide offspring with sufficient milk despite a potential energy deficit. Peak lactation occurs earlier when a fertility index is implemented in the breeding program, allowing for a longer decline in milk yield and hence increasing optimal birth spacing (despite no apparent impact upon the gradient of decay in milk yield).

Correction for environmental effects revealed that high producing dairy cows have a higher genetic potential than is currently supported by the production environment. In contrast, cows with a low fertility and milk yield showed a maximised genetic potential where production was biased upwards by a favourable production environment.

CONCLUSION

Selecting for increased fertility increases initial milk yield and growth rate in amount of daily milk production causing an early production peak. Early peak lactation occurs to provide offspring with sufficient milk despite a potential energy deficit, and provides an increased duration over which milk production declines. This provides sufficient time for the cow to recover from the energy deficit during early lactation prior to a subsequent pregnancy. Further, current production environments could be optimised to fulfil the genetic potential of high producing dairy cows.

ACKNOWLEDGEMENTS

We thank A. Bünger and F. Reinhardt (VIT, Verden, Germany) for providing the EBVs and daughter records.

REFERENCES

Boichard D. and Brochard M. (2012) *Animal* **6:** 544.

Collard B.L., Boettcher P.J., Dekkers J.C., Petitcelrc D. and Schaeffer L.R. (2000) *J. Dairy Sci.* **83:** 2683.

Dekkers J.C.M., Ten Hag J.H. and Weersink a. (1998) *Livest. Prod. Sci.* **53:** 237.

Dekkers J.C. (2007) *J. Anim. Breed. Genet.* **124:** 331.

Ingvartsen K.L., Dewhurst R.J. and Friggens N.C. (2003) *Livest. Prod. Sci.* **83:** 277.

Miglior F., Muir B.L. and Van Doormaal B.J. (2005) *J. Dairy Sci.* **88:** 1255.

Muir B.L., Fatehi J. and Schaeffer L.R. (2004) *J. Dairy Sci.* **87:** 3029.

Oltenacu P.A. and Broom D.M. (2010) *Animal Welfare* **19(S):** 39.

Osteras O., Solbu H., Refsdal A.O., Roalkvam T. Filseth O. and Minsaas A. (2007) *J. Dairy Sci.* **90:** 4483.

Spurlock D.M., Dekkers J.C., Fernando R., Koltes D.A. and Wolc A. (2012) *J. Dairy Sci.* **95:** 5393.

Strucken E.M., Bortfeldt R.H., Tetens J., Thaller G. and Brockmann G.A. (2012) *BMC Genetics* **13:** doi: 10.1186/1471.

Strucken E.M., de Koning D.J., Rahmatalla S.A. and Brockmann G.A. (2011) *J. Dairy Sci.* **94:** 442.

Strucken E.M., Laurenson Y.C. and Brockmann G.A. (2015) *Front. Genet.* **6:** 118.

VanRaden P.M. (2004) *J. Dairy Sci.* **87:** 3125.

Wilmink J.B.M. (1987) *Livest. Prod. Sci.* **16:** 335.

BENCHMARKING COW HEALTH STATUS WITH DAIRY HERD SUMMARY DATA

K.L. Parker Gaddis[1], J.B. Cole[2], J.S. Clay[3], and C. Maltecca[1]

[1] Department of Animal Science, North Carolina State University, Raleigh, NC, USA
[2] Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD, USA
[3] Dairy Records Management Systems, Raleigh, NC, USA

SUMMARY

Genetic improvement of dairy cattle health using producer-recorded data is feasible. Estimates of heritability are typically low, indicating that genetic progress will be slow. Health improvement may also be possible through incorporation of environmental and managerial aspects into herd health programs. The objective of this study was to utilize non-parametric methodologies including support vector machines and random forests to explore prediction of cow health status from routinely collected herd summary data. Random forest models attained the highest accuracy for predicting health status in all health categories. Accuracy of prediction (*SD*) of random forest models ranged from 0.87 (0.06) to 0.93 (0.001). Results of these analyses indicate that non-parametric algorithms, specifically random forest, can be used to accurately identify individual cows likely to experience a health event of interest. Further development of predictive models into herd management programs will continue to improve dairy health.

INTRODUCTION

To fully understand complex diseases, it is important to understand relationships between genotype, environment, and phenotype. Genetic improvement of dairy cattle health has been determined to be feasible utilizing producer-recorded data by several studies (Zwald *et al.* 2004; Parker Gaddis *et al.* 2012, 2014). Low estimates of heritabilities indicate, however, that genetic progress will be slow. Variance observed in lowly heritable traits can largely be attributed to non-genetic or environmental factors. In typical genetic evaluations, adjustments for environmental effects are accomplished by considering them as fixed effects. This disregards potential effects of management and environmental conditions on genetic expression (Windig *et al.* 2005). The question is then whether more rapid phenotypic improvement can be achieved if herd health programs incorporate environmental and managerial aspects.

Recent studies have incorporated herd characteristics into statistical models in relationship to reproductive efficiency (*e.g.*, Löf *et al.* 2007), production (*e.g.*, Windig *et al.* 2006), and health (*e.g.*, Stengärde *et al.* 2012). Farm staff or Dairy Herd Information (*DHI*) Association technicians regularly report on numerous herd characteristics observed on test days (DHI-202: Dairy Records Management Systems 2014). Additional environmental information is accessible through online databases including climatic, human census, and geographical data. Large numbers of variables create analysis challenges, ranging from increased data pre-processing to increased computing time. The majority of previous studies have utilized parametric statistical models to analyse herd characteristics (e.g., Stengärde *et al.* 2012), which can suffer from multiple testing problems and colinearities of numerous variables (Sato *et al.* 2008). Alternatively, non-parametric methods have recently been investigated to better handle numerous variables (*e.g.*, Schefers *et al.* 2010). The objective of this study was to utilize non-

parametric methodologies to explore prediction of cow health status from routinely collected herd summary data.

MATERIALS AND METHODS

Data. The DHI-202 Herd Summary provides a report on herd production, reproduction, genetics, udder health, and feed cost information (www.drms.org). Data were available from 2000 through 2011 from Dairy Records Management Systems (*DRMS*; Raleigh, NC). Four months (March, June, September, and December) of collected records were available for each year. Each herd summary contained over 1,100 variables. Number of contributing herds varied from 647 to 1,418, depending on month and year of reporting. Data included Ayrshire, Brown Swiss, Guernsey, Holstein, Jersey, and crossbred herds.

Supplementary data were acquired from publicly available datasets. The National Oceanic and Atmospheric Administration National Climatic Data Center (*NCDC*) provides information regarding temperatures, precipitation, degree-days, and drought indices (NCDC, 2014). Monthly summaries of data from the weather station located closest to each herd were merged with herd characteristic data. Estimates of population size were obtained on a county-basis from the United States Census Bureau (www.census.gov) as a measure of population density. Intercensal estimates from 2000 through 2010 were produced by updating the Census 2000 counts with estimates for components of population change (United States Census Bureau, 2012).

Voluntary producer-recorded health event data were available from DRMS (Raleigh, NC) from U.S. farms from 2000 through 2012. These data were merged with available production data. Health and production datasets were edited following the editing procedures described in Parker Gaddis *et al.* (2012). Health events included hypocalcemia, cystic ovaries, digestive problems, displaced abomasum, ketosis, mastitis, metritis, and retained placenta. These events were grouped into three main categories: mastitis, metabolic (hypocalcemia, digestive problems, displaced abomasum, and ketosis), and reproductive (cystic ovaries, metritis, and retained placenta) disorders. Health events were combined with herd characteristics based on date of health event occurrence.

Data pre-processing. A function was employed to determine and remove highly correlated variables by searching the correlation matrix. Editing was also performed to ensure that no variables were linear combinations of other variables (Kuhn 2013). Any variables with (near) zero variance were removed from the data. The above editing reduced the size of the dataset to approximately 3.7 million records with 829 variables. Missing records needed to be handled before statistical modeling could be performed. Variables with more than 50% missing observations (n = 70) were excluded from further analyses. Remaining missing herd characteristic records were imputed using an iterative principal component analysis algorithm (Husson and Josse 2012). Once a complete dataset was created, lactational incidence rate was calculated for each health event by herd-year as number of affected lactations per lactations at risk (Kelton *et al.* 1998).

Analyses. Analyses were performed using a binary indicator where “0” represented no incidence of a health event during a lactation and “1” represented at least one incidence of a respective health event during a lactation. Nonparametric models investigated included support vector machines (SVM) and random forests (RF). Briefly, an SVM model maps response variables to a higher-dimensional space that contains a “maximal separating hyperplane” (Sullivan 2012). The response variable should separate across this hyperplane into correct classifications (Sullivan 2012). Two different kernel

functions were investigated: a linear kernel and a radial basis kernel (RBF). The SVM*perf* software (version 3.0) was utilized to fit SVM models (Joachims 2006).

Tree models are a data mining technique that are easily interpretable and implicitly perform feature selection, making them ideal for data with numerous variables (Kuhn and Johnson, 2013). Random forest (RF) models were utilized as a machine learning algorithm that fits many decision trees to bootstrapped samples of a dataset and then averages these decision trees to create a final predictive model (Breiman 2001). The "bigrf" package of R (R Core Team 2014) was used to fit these models (Lim *et al.* 2014). An optimal number of trees was determined prior to fitting a final model by testing a range of values for each health event category.

For all the above described models, 10-fold cross validation was used to evaluate predictive ability. Measures of predictive ability included accuracy, sensitivity, and specificity. Accuracy was calculated as the sum of true positives and true negatives divided by the sum of positive and negative incidences. Sensitivity, or true positive rate, was calculated as number of positive incidences correctly identified divided by the total number of positive incidences. Specificity, or true negative rate, was calculated as the number of negative incidences correctly identified divided by the total number of negative incidences (Fawcett 2006).

RESULTS AND DISCUSSION

The number of states reporting data ranged from 35 to 45, depending on health event. The most common herd size fell in a range of 100 to 299 cows; however, data included herds with fewer than 50 cows and a maximum herd size of over 5,500 cows. Overall median incidence rates were 24%, 8%, and 18% for mastitis, metabolic, and reproductive health events, respectively. These fall within the range of previously reported incidence rates (Parker Gaddis *et al.* 2012).

Predictive ability in training datasets were similar to those estimated for validation data, indicating that the models were not being overfit to training data. Prediction accuracies, sensitivity, and specificity for SVM models are shown in Table 1. Linear and RBF kernels performed similarly for all health event categories. These models had much higher specificity compared to sensitivity, indicating that they were more capable of identifying healthy cows.

Table 1 Summary of model performance for incidences of mastitis, reproductive, and metabolic health events averaged across 10-fold cross validation results fitting support vector machine (SVM) and random forest models

		Accuracy (Validation)	Sensitivity (Validation)	Specificity (Validation)
Mastitis	SVM (linear) c=0.01*	0.70 (0.003)	0.24 (0.002)	0.88 (0.003)
	SVM (RBF) c=10.0	0.70 (0.01)	0.39 (0.03)	0.83 (0.02)
	Random forest	0.93 (0.001)	0.82 (0.003)	0.97 (0.001)
Reproductive	SVM (linear) c=0.005	0.69 (0.002)	0.32 (0.01)	0.79 (0.004)
	SVM (RBF) c=10.0	0.77 (0.01)	0.33 (0.03)	0.88 (0.02)
	Random forest	0.92 (0.001)	0.74 (0.006)	0.97 (0.0007)
Metabolic	SVM (linear) c=0.01	0.76 (0.03)	0.12 (0.03)	0.93 (0.05)
	SVM (RBF) c=10.0	0.75 (0.01)	0.25 (0.02)	0.88 (0.01)
	Random forest	0.87 (0.061)	0.57 (0.145)	0.96 (0.04)

*c represents the error penalty tuning parameter for SVM models

The optimal number of trees for RF models was determined to be 25, regardless of health event. Random forest models had the best predictive ability across all health event categories (Table 1). Overall, sensitivity was lower than specificity; however, sensitivity was higher for RF models compared to SVM models.

Each of the models investigated herein had benefits and disadvantages. Support vector machines are a flexible class of models with several kernels that can be employed. These models require estimation of tuning parameters and results can be more difficult to interpret. Random forests were the most flexible models. They can easily handle a large number of variables, as well as missing records. Random forest models can be more difficult to interpret than a single decision tree, but tend to have better predictive performance and are capable of identifying influential variables.

This study suggests that benchmarking of cow health is feasible with routinely collected data. Improvement in predictive ability may be possible by modeling each health event as opposed to grouping events into categories. Factors that predispose a cow to retained placenta, for example, may not be the same as factors that increase a cow's risk of cystic ovaries. With continued development and incorporation of predictive models into herd management, routinely recorded herd data could be used in conjunction with genomic selection strategies to further improve dairy cattle health.

REFERENCES

Breiman L. (2001) *Mach. Learn.* **45**: 5.

Fawcett T. (2006) *Pattern Recognit. Lett.* **27**: 861.

Husson F. and Josse J. (2012) 'Handling missing values with/in multivariate data analysis (principal component methods).'

Joachims T. (2006) In 'Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining' ACM Press, New York.

Kelton D.F., Lissemore K.D. and Martin R.E. (1998) *J. Dairy Sci.* **81**: 2502.

Kuhn M. (2013) 'Classification and Regression Training.'

Kuhn M. and Johnson K. (2013) 'Applied Predictive Modeling' Springer, New York.

Lim A., Breiman L. and Cutler A. (2014) 'bigrf: Big Random Forests: Classsification and Regression Forests for Large Data Sets.'

Löf E., Gustafsson H. and Emanuelson U. (2007) *J. Dairy Sci.* **90**: 4897.

Parker Gaddis K.L., Cole J.B., Clay J.S. and Maltecca C. (2012) *J. Dairy Sci.* **95**: 5422.

Parker Gaddis K.L., Cole J.B., Clay J.S. and Maltecca C. (2014) *J. Dairy Sci.* **97**: 3190.

R Core Team (2014) 'R: A Language and Environment for Statistical Computing.'

Sato K., Bartlett P.C., Alban L., Agger J.F. and Houe H. (2008) *Acta Vet. Scand.* **50**: 4.

Schefers J.M., Weigel K.A., Rawson C.L., Zwald N.R. and Cook N.B. (2010) *J. Dairy Sci.* **93**: 1459.

Stengärde, L., Hultgren, J., Tråvén, M., Holtenius K. and Emanuelson U. (2012) *Prev. Vet. Med.* **103**: 280.

Sullivan, R. (2012) 'Introduction to Data Mining for the Life Sciences' Springer, New York.

United States Census Bureau. (2012) 'Methodology for the Intercensal Population and Housing Unit Estimates: 2000 to 2010.'

Windig J.J., Calus M.P.L., Beerda B. and Veerkamp R.F. (2006) *J. Dairy Sci.* **89**: 1765.

Windig J.J., Calus M.P.L. and Veerkamp R.F. (2005) *J. Dairy Sci.* **88**: 335.

Zwald N.R., Weigel K.A., Chang Y.M., Welper R.D. and Clay J.S. (2004) *J. Dairy Sci.* **87**: 4287.

PROPOSED CHANGES IN THE GENETIC EVALUATION OF DAIRY FERTILITY IN NEW ZEALAND

K. Stachowicz[1], G.M. Jenkins[1], P.R. Amer[1], J.R. Bryant[2] and S. Meier[2]

[1]AbacusBio Limited, PO Box 5585, Dunedin 9058, New Zealand
[2]DairyNZ Limited, Private Bag 3221, Hamilton 3240, New Zealand

SUMMARY

This paper summarises the findings of a number of research projects exploring the genetic evaluation of dairy cattle fertility in New Zealand and describes the proposed changes to the genetic evaluation system. The following recommendations for changes in the national genetic evaluation model of fertility traits were made: 1) replace the binary calving trait with its continuous version measured in days; 2) include records on first calving of heifers in the model; 3) include the information on reproductive interventions like labour inductions and mating hormonal synchronization for evaluating fertility phenotypes; 4) drop milk production traits from the model; 5) retain body condition score as a predictor trait in the model; 6) add gestation length as a new correlated predictor trait to the model; 7) consider the inclusion of a gestation length breeding value (direct) in selection index.

INTRODUCTION

The reproductive performance of dairy cows is a major driver of profit in dairy farming, especially in seasonal, pasture-based systems, where the calving interval is essentially constrained to a short window around 365 days. Herds that achieve high levels of fertility (i.e. meet the industry target of 78% 6-week in-calf rate with minimal use of hormonal intervention) are able to maintain a condensed seasonal calving pattern and minimise the number of cows being culled involuntarily for reproductive reasons. By international standards, reproductive performance of dairy cows in New Zealand is high. It has, however, experienced the global decline in fertility that was associated with selection for high milk yields without the protection of including fertility in the economic selection index, known as Breeding Worth (BW; Harris and Montgomerie 2001).

In New Zealand, fertility is currently evaluated using two binary traits: calving rate at 42 days (CR42; if cow calves within 42 days from the planned start of calving in parities 2 to 4) and percentage mated at 21 days (PM21; if the cow is presented for mating within 21 days from planned start of mating) in parities 1-3, along with two predictor traits: first lactation milk yield and body condition score (BCS). These traits indirectly measure a cow's genetic propensity to return to a fertile state after calving and to become pregnant in a period that will allow her to maintain a 365-day calving interval. Both PM21 and CR42 have low heritability (0.05 and 0.02, respectively; Harris *et al.* 2005).

There are potential benefits of incorporating new fertility traits into the genetic evaluation, particularly if these traits are more accurate measures of true fertility or if they can be measured earlier in the cow's life than current mating and calving traits. A number of research projects were undertaken in order to explore possible improvements to the national genetic evaluation system for fertility.

Bowley *et al.* (2015) analysed data from 169 herds participating in fertility monitoring projects (Brownlie *et al.* 2011) in the 2011 and 2012 seasons and demonstrated that redefining fertility as a continuous calving trait (CSD - calving season day, which is defined as the difference in days between planned start of calving for a given contemporary group and actual cow's calving date) instead of using binary (0/1) scores as per the current trait definition could increase the accuracy of fertility evaluation. A heifer calving trait was also investigated and shown to have similar

heritability estimates to cow fertility traits. In addition, there was a relatively high genetic correlation between heifer and cow fertility (0.7) which means that some of the genes underlying cow fertility are similar to those underlying heifer fertility. Hence, the heifer fertility trait may be useful in predicting the estimated breeding value (EBV) for fertility of cows and young bulls (Pryce *et al.* 2007). Adjusting for reproductive interventions by setting affected calving and mating records to missing was also suggested (Bowley *et al.,* 2015). Novel potential fertility traits were investigated and postpartum anoestrous interval was recommended to be considered as a trait with heritability higher than currently evaluated traits (0.08). However, its inclusion on a national scale would require a substantial effort by the industry to record phenotypic data.

Stachowicz *et al*. (2014a) evaluated the initial finding by Bowley *et al.* (2015) using national fertility data. Firstly, this research addressed the issue of the impact of alternative approaches of identifying planned start of mating and planned start of calving dates for each contemporary group on heritability estimates of fertility traits and found minimal differences. Secondly, the genetic parameters PM21 and CR42 were estimated using alternative trait definitions (continuous vs binary vs a scoring system) to determine the best fertility traits for subsequent work. From this, it was recommended that the binary version of the mating trait and the continuous version of the calving traits should be used. Thirdly, genetic correlations of the best fertility traits with other fertility, production and conformation traits were estimated to identify possible predictor traits for fertility. Results indicated that milk yield, the current predictor trait, might be better replaced with protein percentage. Finally, the heritability of heifer calving season day (CSD) was estimated, as well as its genetic correlations with cow fertility traits and this was found to be a valuable early predictor trait of future cow fertility.

Selection index modelling methods were used to investigate the impact of changes in the fertility genetic evaluation model on accuracy of estimated breeding values for fertility traits (Stachowicz *et al.*, 2014b). The binary or continuous equivalents of the current fertility traits were compared, replacing milk yield with protein yield or protein percent as a predictor trait in the fertility model and by considering the impact of inclusion of the heifer fertility trait. The results of this study agreed with previous findings, that: 1) the continuous version of CR42 was a more accurate predictor of true fertility than the currently used binary trait; 2) including milk production traits in the fertility model has minimal impact on accuracy of the prediction for fertility; and, 3) adding the heifer calving trait increased the accuracy of cow fertility trait predictions.

The research reported here continues the work described by Stachowicz *et al.* (2014a & 2014b). Gestation length (GL) expressed as trait of the calf was investigated as a possible predictor trait to be included in the genetic evaluation model for cow fertility. Variance components were estimated and the selection index model was extended to include this trait in order to evaluate the impact it would have on the accuracy of the fertility breeding value prediction.

MATERIALS AND METHODS

Data. The fertility data were extracted from the New Zealand national dairy database. This data included records from 1989 to 2013 for cows having at least heifer calving recorded. Mating and calving records from the first three lactations were considered. Data edits and fertility trait (PM21 and CSD) definitions were described in detail by Stachowicz *et al*. (2014a). Gestation length data were included for the cows that had fertility phenotypes, as well as for all the calves born to those cows. Gestation length was calculated as a difference in days between last recorded mating and calving date. Only artificial insemination matings were included.

Genetic analysis. Genetic analysis was carried out using ASReml software (Gilmour *et al.*, 2009). The fertility traits (PM21 and CSD) were analysed with the same model which corresponds to the one used in national genetic evaluations of fertility traits (Harris *et al.* 2005):

$$PM21 \text{ or } CSD = CG + Age*Breed + HF_{FR} + HF_{NZ}$$

$+ HET_{JExNZ} + HET_{JExFR} + HET_{NZxFR} + REC_{JExNZ} + REC_{JExFR} + REC_{NZxFR} + a + e.$

Gestation length was analysed with the following direct- maternal model:

$$GL = CG + SireCode + AgeD*BreedD + SexC*BreedC + dHF_{FR} + dHF_{NZ} + cHF_{FR} + cHF_{NZ}$$
$$+ dHET_{JExNZ} + dHET_{JExFR} + dHET_{NZxFR} + dREC_{JExNZ} + dREC_{JExFR} + dREC_{NZxFR}$$
$$+ cHET_{JExNZ} + cHET_{JExFR} + cHET_{NZxFR} + cREC_{JExNZ} + cREC_{JExFR} + cREC_{NZxFR}$$
$$+ a + m + mpe + e,$$

where:

- *CG* is the fixed contemporary group effect (herd-year-age for PM21 and CSD; herd-year-age of the dam for GL)
- *SireCode* is the sire official indicator
- *Age*Breed* is the fixed linear regression of age at calving in days nested within breed (for heifers the Julian day of the year when the cow was born was used)
- *AgeD*BreedD* is the fixed linear regression of age of the dam in days nested within breed of the dam
- SexC**BreedC* is the fixed linear regression of sex of the calf nested within breed of the calf
- HF_{NZ} and HF_{FR} are fixed linear regressions of New Zealand and foreign Holstein-Friesian breed composition (*d* for dam; *c* for calf)
- *HET* and *REC* are fixed linear regressions of breed specific heterosis and recombination effects (*d* for dam; *c* for calf)
- *a* is a random animal effect
- *m* is a random maternal effect (model allowed for covariance between *a* and *m*)
- *mpe* is a random maternal permanent environmental effect
- *e* is a random error term.

The multivariate animal model included 8 traits: PM21 in parities 1-3, CSD in parities 0-3 (where 0 denotes heifer calving), and GL.

Selection index modelling. The selection index model, as described by Stachowicz *et al.* (2014b), was extended to predict the effects of including GL as predictor a trait in the fertility genetic evaluation model. It was also used to assess the effects of placing direct selection emphasis on GL (by including it in the BW index) on the rates of genetic change that might be expected in GL, fertility, and the overall annualised profit from dairy cattle genetic improvement in New Zealand. Direct selection emphasis on GL, in addition to using it as a correlated predictor of daughter fertility, may be justified in the future based on the effect the bull has as a service sire on the fertility of the cows to which he is mated.

RESULTS AND DISCUSSION

Genetic parameters. Estimated heritabilities, as well as genetic and residual correlations are presented in Table 1. Gestation length is a trait with very high heritability (0.68) as well as high and favourable genetic correlations with evaluated fertility traits (PM21 and CSD), while corresponding residual correlations remain low, which makes GL a perfect candidate as a predictor trait for fertility traits in a multiple trait evaluation.

Selection index modelling. These results indicate that including direct GL as a predictor trait in the fertility genetic evaluation model, in addition to the changes already considered by Stachowicz *et al.* (2014b), would further increase the accuracy of the fertility breeding value prediction. If GL was to be utilized in the BW index as well as a predictor trait in the fertility genetic evaluation model, then a further increase in the rate of genetic gain would be expected for both female fertility and GL. Including GL as a predictor trait in the fertility model and the BW index could increase annual industry profit by about $7 million. Including GL in the BW index only would allow for capturing around 20% of this amount. The phenotypic effects of GL on future fertility, production, survival and other traits were investigated and discussed by Jenkins *et*

al. (2015).

Table 1. Residual correlations (below diagonal), genetic correlations (above diagonal) and heritabilities (diagonal) for PM21 in parities 1-3, CSD in parities 0-3 and GL (A=direct animal effect, M=maternal effect)

	CSD_0	CSD_1	CSD_2	CSD_3	PM21_1	PM21_2	PM21_3	GL_A	GL_M
CSD_0	**0.02**	0.66	0.39	0.40	0.07	0.22	0.04	0.57	0.08
CSD_1	0.12	**0.02**	0.30	0.34	-0.25	-0.03	-0.06	0.56	0.09
CSD_2	0.00	0.34	**0.01**	0.70	-0.49	-0.58	-0.42	0.45	0.09
CSD_3	0.03	0.06	0.16	**0.02**	-0.50	-0.47	-0.58	0.38	0.09
PM21_1	-0.12	-0.17	-0.05	-0.01	**0.03**	0.54	0.52	-0.15	-0.07
PM21_2	-0.04	-0.08	-0.16	-0.07	0.10	**0.06**	0.77	-0.21	-0.06
PM21_3	-0.03	-0.06	-0.11	-0.27	0.07	0.14	**0.05**	-0.22	-0.07
GL	0.01	-0.02	0.07	0.07	0.08	0.04	-0.02	**0.68**	-0.26

Recommendations. Based on data from a number of research projects, the following recommendations for changes in the national genetic evaluation model of fertility traits are made: 1) replace the binary CR42 with its continuous version CSD; 2) include heifer CSD in the model; 3) include the information on reproductive interventions in fertility phenotypes; 4) drop milk production traits from the model; 5) retain BCS in the model; 6) add GL as a new predictor trait to the model; 7) consider the inclusion of a direct GL EBV in the BW index to account for the advantage of shorter GL sires to the cows they are mated, over and above the improvements in the fertility of their daughters from shorter GL.

Future work. Variance components for the recommended model are currently being estimated. The next steps include the implementation of a new model for genetic evaluation to be run using the full national fertility data set. New results will be compared with the current genetic evaluation outcomes, prior to releasing new fertility breeding values to industry for consultation.

ACKNOWLEDGEMENTS

This project is funded by New Zealand dairy farmers through DairyNZ Incorporated and the Ministry for Primary Industries through the Primary Growth Partnership.

REFERENCES

Bowley F.E., Green R.E., Amer P.R. and Meier S. (2015) *J. Dairy Sci.* **98**: 2005.

Brownlie T.S., Weir A.M, Tarbotton I., Morton J.M., Heuer C. and McDougall S. (2011) *N. Z. Vet. J.* **59**: 28.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsni.co.uk.

Harris B.L. and Montgomerie W.A. (2001) *Interbull Bull.* **27:** 139.

Harris B.L., Pryce J.E., Xu Z.Z. and Montgomerie W.A. (2005) *Interbull Bull.* **33**: 47.

Jenkins G.M., Amer P., Stachowicz K. and Meier S. (2015) *J. Dairy Sci.* Submitted.

Pryce J.E., Harris B.L. and McNaughton L.R. (2007) *Proc. NZ Soc. Anim. Prod.* **67**: 388.

Stachowicz K., Jenkins G.M., Amer P.R. and Meier S. (2014a) *Interbull Bull.* **48**: 89.

Stachowicz K., Jenkins G.M., Dennis N.A., Amer P.R., Bryant J. and Meier S. (2014b) *Proc. of the 5th Australasian Dairy Science Symposium*, pp 65.

USING SEQUENCE DATA TO IMPROVE ACCURACY OF GENOMIC PREDICTION AND QTL DISCOVERY FOR DAIRY COW FERTILITY

I.M. MacLeod[1,2,3], B.J. Hayes[2,3,4], M. Haile-Mariam[2,3], P. Bowman[2,3], A.J.Chamberlain[2,3], C.J Vander Jagt[2,3], C. Schrooten[5] and M.E. Goddard[1,2,3]

[1] Faculty of Veterinary & Agricultural Science, University of Melbourne, Victoria, Australia.
[2] Dairy Futures Cooperative Research Centre, AgriBio, Bundoora Victoria, Australia.
[3] AgriBio, Dept. Economic Dev., Jobs, Transport & Resources. Victoria, Australia.
[4] Biosciences Research Centre, La Trobe University, Victoria, Australia.
[5] CRV, 6800 AL, Arnhem, Netherlands

SUMMARY

Using a Bayesian genomic prediction method, BayesR, we demonstrate improved accuracy of genomic prediction for cow fertility using high density SNP markers combined with imputed sequence variants in and close to gene coding regions. We also used the same analysis to identify candidate genes and potential causal mutations with a broad range of effects on fertility.

INTRODUCTION

Dairy cow fertility has caused much concern over the past three decades in many countries because it had been in decline, partly due to an unfavourable genetic correlation between fertility and milk traits. Accurate genomic predictions of fertility would be of great benefit to the dairy industry because it is measured only in mature females and has low heritability. It is also important to identify genes affecting fertility to better understand genetic factors that underpin the trait.

BayesR, a Bayesian genomic prediction method, can achieve higher accuracy of genomic prediction compared to genomic best linear unbiased prediction (GBLUP), particularly for traits affected by many small effect genes as well as some of much larger effect (Erbe *et al.* 2012, Kemper *et al*. 2015). This occurs because BayesR models the single nucleotide polymorphism (SNP) effects as a mixture of four normal distributions, including a null distribution and one distribution with moderate to large variance.

BayesR should also be a more precise method for QTL discovery than GWAS (genome-wide association analysis) or GBLUP. GWAS fits SNP individually which often results in one QTL being predicted by a large number of SNP in LD. GBLUP fits all SNP simultaneously but effects are distributed as a single normal distribution so are smeared across many adjacent SNP with strong shrinkage of larger effects. Furthermore, BayesR provides a well calibrated test of the likelihood that a SNP predicts a real QTL effect (posterior probability).

Using BayesR we compare accuracy of genomic prediction for dairy cow fertility using high density SNP markers and imputed sequence variants in and close to genes. We also identify candidate genes and potential causal mutations associated with fertility.

MATERIALS AND METHODS

We obtained dairy bull progeny test phenotypes of female fertility for 6804 bulls, including 5285 black and white Holsteins, 620 Red Holsteins, 803 Jerseys and 96 Australian Reds. Most bulls had MACE international breeding values and these were converted to de-regressed proofs (DRP) on the Australian scale (ie. corrected phenotypes: details in Haile-Mariam *et al.* (2015) – we used a subset of their data). The remainder (252) had daughter trait deviations (DTD) from the Australian Dairy Herd Improvement Scheme (ADHIS). Both the DTD and DRP were converted to the same scale using linear regression. The 620 Red Holsteins were our validation set and the remaining bulls (6184) made up the reference set.

All bulls were either genotyped or imputed for the Bovine HD SNP Illumina array ("800K") as described in Haile-Mariam *et al.* (2015). Bulls were then imputed for sequence variants in coding regions and in "regulatory regions" (defined as 5000bp either side of genes) using Beagle3 (Browning and Browning 2009). Run 3 of the 1000 Bull Genomes Project was used to discover these variants and provided the reference Holstein and Jersey bulls for imputation (Daetwyler *et al.* 2014). The 800K and imputed sequence genotypes were combined to give a third genotype set ("SEQ"). Very rare variants were pruned from SEQ (minor allele frequency: MAF < 0.002) as well one of any pair of SNP in perfect LD, preferentially keeping non-synonymous coding variants and then variants in regulatory regions. A total of 907,643 SEQ variants remained, most of which were SNP and a small number of indels.

BayesR was implemented as detailed in Kemper *et al*. (2015). Briefly we fitted the model:

$\mathbf{y} = \mathbf{Xb} + \mathbf{Wv} + \mathbf{Za} + \mathbf{e}$, where $\mathbf{y}$=vector of phenotypes, $\mathbf{X}$=fixed effects design matrix and $\mathbf{b}$=vector of fixed effects (mean, breed, data type - DRP/DTD nested within breed). Here $\mathbf{W}$=design matrix of variant genotypes centred and standardized (Yang *et al.* 2010) and $\mathbf{v}$=vector of variant effects, distributed as a mixture of four distributions: $N(0, 0.0\sigma^2_g)$, $N(0, 0.0001\sigma^2_g)$, $N(0, 0.001\sigma^2_g)$, $N(0, 0.01\sigma^2_g)$ where σ^2_g=total genetic variance. $\mathbf{Z}$=polygenic effects design matrix, $\mathbf{a}$=vector of random polygenic effects ~ $N(0, \mathbf{A}\sigma^2_a)$, where $\mathbf{A}$=pedigree relationship matrix and σ^2_a=additive genetic variance not explained by the genotypes. $\mathbf{e}$=vector of residual errors ~ $N(0, \mathbf{E}\sigma^2_e)$ where $\mathbf{E}$=diagonal matrix with a weighting coefficient based on effective number of daughter records per bull (Garrick *et al.* 2009). Analyses were performed with 50K, 800K or SEQ genotypes, each running for 40,000 iterations (20,000 discarded as burn-in) with 5 replicates per analysis. Results were derived from the mean of 20,000 iterations, averaged over all replicates. The posterior probability of a given SNP being included in the model was calculated as the proportion of iterations (post burn-in) that each variant fell in a non-zero distribution (averaged across replicates). Accuracy of genomic prediction was estimated as the correlation between predicted breeding value and corrected phenotypes. The mean squared error of the prediction for each analysis was calculated as the average of the squared difference between predicted breeding value and phenotype.

RESULTS AND DISCUSSION

There was a clear trend for the accuracy of genomic prediction to increase with the density of genotypes (Table 1). The highest accuracy was achieved using SEQ genotypes, possibly because some causal mutations were included in the SEQ data. Also the imputed sequence included many more rare variants than in the 50K or 800K data (Table 1). Therefore, SEQ is more likely to include variants in strong LD with other rare/recent mutations, including rare causal mutations, than common SNP on the 50K or 800K arrays.

Table 1. Genomic prediction using different densities of genotypes

Genotype Sets (Total number variants)	**50K** (37,236)	**800K** (600,640)	**SEQ** (907,643)
Proportion of variants with MAF[1] < 0.05	11%	7%	22%
Accuracy of Genomic Prediction (s.e.m.[2])	**0.386** (0.0008)	**0.418** (0.0004)	**0.440** (0.0012)
Mean Squared Error of prediction	166	148	144

[1] Minor Allele Frequency

[2] Calculated as: SD of accuracy from 5 replicates divided by √5

To better understand the contribution of sequence variants in and near coding regions, we tested the accuracy of prediction in the validation set using only the top 5000 variants (based on posterior probabilities) from either:

A. Non-synonymous coding variants and variants within 5000bp of genes, or
B. Intergenic variants excluding regions close to genes (±5000bp).

Prediction accuracy for group A = 0.35 while B = 0.27. When sets A and B were combined, prediction accuracy = 0.39. The high LD in the cattle genome means that there is considerable overlap in the predictions from set A and B. However, the results do suggest that variants in and close to coding regions explain a large proportion of the trait variance but that intergenic regions are also important in regulating trait expression, in keeping with evidence from the human ENCODE project (Skipper *et al.* 2012). The fertility trait appeared to be highly polygenic, with an average of 3305 SNP effects fitted in the model. This is expected because the fertility trait was largely based on calving interval: a complex trait influenced by many factors such as cow energy balance, oocyte health and embryo development.

In Figure 1 we present some examples of QTL discovery from among the 50 most significant QTL regions occurring within 5000bp of a gene: first to demonstrate the advantages of our approach and second to illustrate the range of genetic factors that underpin the complex female fertility trait.

Several regions only showed strong evidence for QTL in the SEQ analysis, demonstrating the improved power of SEQ genotypes. One example (Fig. 1a) is a rare variant in the 3' UTR of the SCARA5 gene showing a strong probability of being either the causal variant or one in strong LD with a QTL. This SNP is likely a relatively recent mutation because it segregated only in the black and white Holsteins (MAF=0.08) and was not in strong LD with any other SNP. SCARA5 expression is upregulated in human endometrium tissue when an early embryo is present (Duncan *et al.* 2011), and was also found to be more highly expressed in bovine ovary tissue compared to 17 other tissues (Chamberlain *et al*. 2014). A second example (Fig 1b) is two SEQ variants in high LD (Holstein only, MAF=0.025). The highest probability variant lies between SMEK1 and CCDC88C gene, while the other is a missense mutation in SMEK1. Potentially either gene could be considered to be a good candidate. CCDC88C is a negative regulator in the Wnt signalling pathway that regulates embryo germ cell development (Enomoto *et al.* 2006). SMEK1 has been demonstrated to regulate hepatic gluconeogenesis in mice (Yoon *et al.* 2010) and also appears to regulate the differentiation of embryonic stem cells (Lyu *et al*. 2011). In an analysis of the same data for milk traits (results not shown), these same mutations have a strong association with milk yield and the allele that increased milk yield reduced fertility.

A further region that showed a strong association for fertility and milk yield is between the GC and NPFFR2 genes (Fig. 1c). There is strong LD across this region and the association was spread across several variants. Again both genes are potentially good candidates: GC encodes Vitamin D transporter and disruption of the Vitamin D pathway affects oestrogen biosynthesis, while NPFFR2 interacts with kisspeptin which plays a key role in neuroendocrine regulation of reproduction (Matzuk and Lamb 2008). A number of regions on the X chromosome showed several strong QTL signals including SNP very close to KAL1 (Fig 1d) and UBE2A. In humans several mutations in KAL1 are responsible for “Kallmann syndrome”, affecting the embryonic migratory pathway of neurons that synthesize gonadotropin-releasing hormone (Hardelin *et al.* 1992). This results in impaired gonad development in males and females. Mutations in UBE2A have been shown to be associated with maternal effects on early embryo survival (Matzuk and Lamb 2008).

The validity of our results are dependent on the accuracy of imputation and reference genome annotation, neither of which is perfect. However, this study demonstrates that imputed sequence genotypes with Bayesian analysis improved the accuracy of genomic prediction and the QTL discovery highlighted a broad range of genetic factors potentially affecting dairy cow fertility.

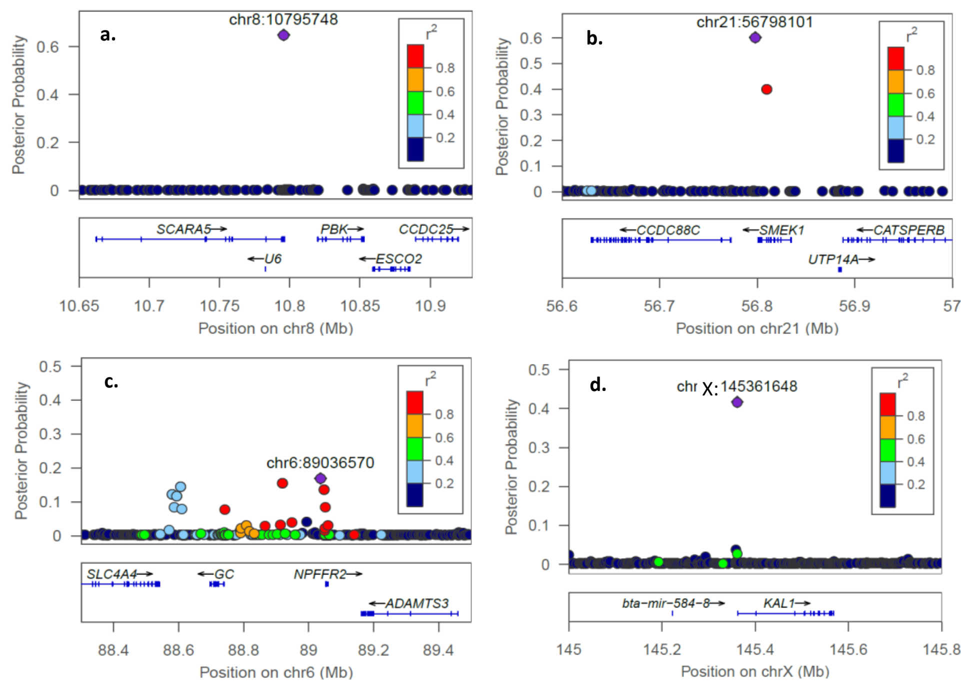


Figure 1. QTL discovery: The variant with the highest BayesR posterior probability is annotated and shown as purple diamond, and the LD (r^2) between this and other variants is colour coded. Genes are shown in blue with exons delineated by thicker bars.

REFERENCES

Browning B.L. and Browning S.R. (2009) *American J Human Genetics* **84**: 210.

Chamberlain A.E., Vander Jagt C.J., Hayes B.J. and M. Goddard M.E. (2014) *Proc 10th World Cong. Genet. Appl. Livestock Prod*. Paper 180.

Daetwyler H.D., Capitan A., Pausch H., Stothard P., Van Binsbergen, R., *et al*. (2014) *Nat Genet.* **46**: 858.

Duncan W.C., Shaw J.L.V., Burgess S., McDonald S.E., Critchley H.O.D. and Horne A.W. (2011) *PLoS ONE* **6**: e23595.

Enomoto A., Ping J. and Takahashi M. (2006) *Ann. N Y Acad. Sci.* **1086**: 169.

Erbe M., Hayes B.J., Matukumalli L.K., Goswami S., Bowman P.J., *et al*. (2012) *J. Dairy Sci.* **95:** 4114.

Garrick D.J., Taylor J.F. and Fernando R.L. (2009) *Genet. Sel. Evol.* **41:** 44.

Haile-Mariam M., Pryce J.E., Schrooten C. and Hayes B.J. (2015) *J. Dairy Sci*. **98**: 3443

Hardelin J.P., Levilliers J., del Castillo I., Cohen-Salmon M., Legouis R., *et al*. (1992) *Proc. Nat. Acad. Sci.* **89:** 8190.

Kemper K., Reich C., Bowman P., Vander Jagt C., Chamberlain A., *et al*. (2015) *Gen. Sel. Evol.* **47**: 29.

Lyu J., Jho E.-h. and Lu W. (2011) *Cell research* **21**: 911.

Matzuk M.M. and Lamb D.J. (2008) *Nat. Med.* **14:** 1197.

Skipper M., Dhand R. and Campbell P. (2012) *Nature* **489**:45.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., *et al*. (2010) *Nat. Gen.* **42**: 565.

Yoon Y.-S., Lee M.-W., Ryu D., Kim J.H., Ma H., *et al*. (2010) *Proc. Nat. Acad. Sci.* **107:** 17704.

GENOMIC BREEDING VALUES OF HEAT TOLERANCE IN AUSTRALIAN DAIRY CATTLE

T.T.T. Nguyen[1], P.J. Bowman[1], M. Haile-Mariam[1], J.E. Pryce[1,2] and B.J. Hayes[1,2]

[1]BioSciences Research Division, Department of Economic Development, Jobs, Transport and Resources, 5 Ring Road, Bundoora, Victoria 3083, Australia
[2]La Trobe University, Bundoora, Victoria 3083, Australia

SUMMARY

In this study, we aimed to develop genomic estimated breeding values for heat tolerance in Australian dairy cattle. We combined test-day herd recording data with temperature and humidity measurements (in the form of temperature-humidity index or THI) from weather stations that were closest to the herds for test days between 2003 and 2013. Tolerance to heat stress was then estimated for each cow using random regression (intercept and slope) to model the rate of decline in production with increasing THI accumulated over the four days prior to the day of milking, for milk, fat and protein yields. The cow slopes from this model were used to define daughter trait deviations (DTD) for their sires. Data were analysed separately for Holsteins and Jerseys. The reference population for genomic prediction was 2,300 Holstein and 575 Jersey genotyped sires with DTD for response to heat stress for milk, fat and protein yield. With this reference, and using GBLUP, the range in accuracy of genomic predictions for heat tolerance across traits were 0.38 – 0.53 and 0.49 – 0.63 for 435 Holstein and 135 Jersey validation sires, respectively. When 2,191 Holstein and 1,190 Jersey cows were added in the reference populations, no substantial improvements in accuracy were observed. Genomic selection appears to be a useful tool to enable farmers to improve milk production in environments with higher heat load.

INTRODUCTION

Changes in environmental factors such as air temperature, humidity, air flow and radiation beyond the comfort zone of animals will lead to heat stress (Armstrong 1994). Heat stress in dairy cattle is an important issue as it results in reduced milk yield (Hayes *et al.* 2003), reduced fertility (Haile-Mariam *et al.* 2008) and therefore reduced profitability (St-Pierre *et al.* 2003). As the temperature in Australia is projected to continue to increase, the future of dairy farming will need measures to adapt to heat stress.

One way to address the challenge posed by heat stress is to apply management measures such as providing shade, fans and sprinklers to cows. Another approach that may have greater benefits in the long term is to select animals with better heat tolerance, as it has been demonstrated that variation in heat stress response is heritable (Hayes *et al.* 2003; Haile-Mariam *et al*. 2008). That is, production and fertility of some animals are less affected by heat stress than others and therefore these animals could be valuable candidates for selection. The genetic gain for heat tolerance will be greatest if accurate genomic estimated breeding values are available, as this will enable selection of young bulls and heifers.

In this study, we used dense DNA markers, together with field production and climate data, to develop GEBV for heat tolerance for dairy cattle in Australia.

MATERIALS AND METHODS

Hourly climate data including dry bulb temperature and relative humidity (%) were obtained from the Bureau of Meteorology (Melbourne, Australia) for all weather stations in Australia from 2003 to 2013. Average temperature-humidity index (THI) on the test day and 1, 2, 3 and 4 d

before the test day were calculated following Hayes *et al.* (2009). The first-lactation test-day records between 2003 and 2013 for milk, protein and fat yield were extracted from the Australian Dairy Herd Improvement Scheme (ADHIS) database for Holstein and Jersey cows. Production records were merged with THI from the nearest weather station, or the second nearest station if it had a lower number of missing records. In total, THI from 105 weather stations were matched to production records of 1,655 Holstein and 501 Jersey dairy herds.

In our dataset, genotypes were available for a total of 2,735 Holstein and 710 Jersey sires. Illumina Bovine High-Density genotypes (777,963 SNP markers) were available for 1,620 of the Holstein sires and 125 of the Jersey sires. For all other sires, 50K (56,430 SNP) genotypes were available. After quality control and removal of non-polymorphic SNPs, 632,004 SNPs remained for animals genotyped at high density and 43,425 SNP remained for animals genotyped at the lower density. All animals genotyped at the lower density had genotypes imputed to the higher density SNP panel using BEAGLE 3 (Browning and Browning 2009).

All statistical analyses were undertaken separately for Holstein and Jersey. Mixed linear models were used to fit the data with variance components estimated using maximum likelihood in ASReml (Gilmour *et al.* 2009).

A random regression model was used to derive individual cow sensitivity to changes in THI of milk, fat and protein yields (i.e. the slope of the regression, or cow slope): $y_{ijl} = \mu + HTD_i + YS_j + \sum_{n=1}^{3} A_n X_n + \sum_{n=1}^{8} D_n Z_n + \sum_{n=0}^{l} P_{ln} W_n + e_{jli}$ (model 1), where y_{ijl} is yield of milk in litres, fat in kg x100 or protein in kg x 100 from the ith herd test day, j^{th} year season of calving, and l^{th} cow in her first lactation; μ is the intercept, HTD_i is the effect of the ith herd test day; YS_j is the effect of the jth year season of calving; X_n is the nth-order orthogonal polynomial corresponding to age on day of test; A_n is a fixed regression coefficient of milk/fat/protein yield on age at test; Z_n is the nth-order orthogonal polynomial corresponding to days in milk (DIM) at test; D_n is a fixed regression coefficient of milk/fat/protein yield on DIM; P_{ln} is a random regression coefficient on THI for the lth cow; W_n is either the intercept or slope solution for heat load index for cows; and e_{ijl} is the vector of residual effects. In this random regression model, all THI values below 60 were set to 60 (Hayes *et al.* 2009).

The effects of the sires (sire slope) to sensitivity of milk, protein and fat yield of cows to changes in THI were obtained using the following model: $y_i = \mu + Sire_i + e$ (model 2), where, y_i is a vector of slope value for a daughter of the ith sire obtained from model 1, $Sire_i$ is the effect of the ith sire on cow slope ~ $N(0, \sigma_s^2)$, e is the vector of residuals ~ $N(0, \boldsymbol{I}\sigma_e^2)$ where $\boldsymbol{I}$ is identity matrix and σ_e^2 is residual variance.

Proportion of additive variance in cow slope was calculated as 4 times of sire variance divided by total variance obtained from model 2.

To assess the accuracy of using genomic breeding values to predict heat tolerance, in each breed, the sires were split into a reference and a validation population. These populations were split by age, with sires born before 2005 included in the reference population, and sires born in or after 2005 placed in the validation population for Holsteins; sires born before 2004 were included the reference population, and sires born in or after 2004 were placed in the validation population for Jerseys. Sires that are paternal half-sibs were placed in either the reference set or the validation set. The genomic breeding values (GEBV) of the validation sires (the phenotype of which were not included in the analysis), were estimated by GBLUP using model: $y = \mu + Zg + e$ (model 3), where y is a vector of sire slope values (solutions for model 2), μ is the intercept, Z is a design matrix relating records to genomic breeding value of animals, g is a vector of genomic breeding values ~ $N(0, \boldsymbol{GRM}\sigma_g^2)$, where $\boldsymbol{GRM}$ is the genomic relationship matrix, σ_g^2 is the additive genetic variance captured by SNP, and e is a vector of random residuals ~ $N(0, \boldsymbol{W}\sigma_e^2)$, where $\boldsymbol{W}$ is identity

matrix and σ_e^2 is residual variance. The genomic relationship matrix (***GRM***) was constructed amongst all genotyped individuals following Yang *et al.* (2010).

Proportion of additive variance of sire slope that is explained by SNP was calculated as additive variance divided by total variance obtained from model 3.

In some analyses the genotyped cows were used in the reference population as well as reference sires, and in this case cows that were daughters of validation bulls were excluded from the analyses. A similar model to model 3 was fitted to the reference data, but in this case the difference in residual variances for bull and cow phenotypes were taken into account by constructing the diagonal matrix $\boldsymbol{W}$ as $g(1/w_i)$, where w_i is the weighting coefficient for the ith animal. Weighting coefficient was calculated differently for bulls and cows following Garrick *et al.* (2009), as follows:

$$w_i(bulls) = \frac{1-h^2}{ch^2+\frac{4-h^2}{p}} \quad ; \quad w_i(cows) = \frac{1-h^2}{ch^2+\left[\frac{1+(n-1)t}{n}-h^2\right]}$$

where $h^2$is the heritability of heat tolerance, c is the proportion of variance not explained by SNP (c =0.2), d is the effective number of daughters, n is the number of repeat records (i.e. lactations), r is the number of records per cow, and t is the repeatability of the trait (average repeatability for cow slopes in relation to milk, fat and protein was 0.34 for Holsteins and 0.44 for Jerseys in the current datasets).

The accuracy of genomic prediction was taken as the correlation of the genomic estimated breeding values, and the slopes for the validation sires (from model 2), divided by square root of the proportion of variance of sire slope explained by SNP obtained from model 3.

RESULTS AND DISCUSSION

Table 1. Correlations between sire slopes and GEBV, and accuracies of genomic estimated breeding values using 632,004 SNP panel for heat tolerance on milk, fat and protein yield using GBLUP

Breed	Reference	Validation	Production traits affected by heat stress	Correlation	Accuracy
Holstein	2,300 sires	435 sires	Milk yield	0.26	0.46
			Fat yield	0.22	0.38
			Protein yield	0.26	0.47
	2,300 sires + 2,191 cows	435 sires	Milk yield	0.27	0.48
			Fat yield	0.22	0.38
			Protein yield	0.29	0.53
Jersey	575 sires	135 sires	Milk yield	0.34	0.49
			Fat yield	0.40	0.60
			Protein yield	0.42	0.63
	575 sires + 1,190 cows	135 sires	Milk yield	0.39	0.56
			Fat yield	0.39	0.60
			Protein yield	0.43	0.64

There was substantial test-day milk, fat and protein yield variation in the datasets. In all dairy farming regions of Australia, cows experienced some degree of heat stress, which was variable among regions. For example, in Queensland the median of daily temperatures and **THI** were 20.7 and 67.3, respectively. Northern Victoria experienced a wide spectrum of weather conditions, with

average daily temperatures ranging from -5°C to 38°C (median of 13°C), and **THI** ranging from 33.8 to 81.0 (median 56.6).

The proportion of additive variance in cow slope in relation to milk, fat and protein yields were 0.14, 0.16 and 0.13 in Holsteins and 0.27, 0.26 and 0.21 in Jerseys, respectively. The corresponding proportions of variance explained by SNP in sire slope were 0.32, 0.34 and 0.30 in Holsteins and 0.49, 0.44 and 0.45 in Jerseys. This confirms that selection for heat tolerance is possible and could be particularly effective in the geographical regions with high heat load. Correlations of sire slopes among milk, fat and protein yield ranged from 0.46 – 0.86 for Holsteins and 0.60 – 0.86 for Jerseys.

Genetic correlations between heat tolerance and production traits in cows were negative. For example, in Holsteins the correlations between heat tolerance with milk, fat and protein yields were -0.38, -0.40 and -0.54, respectively. This confirms the antagonistic relationships between heat tolerance and production traits reported in previous studies (Ravagnolo and Misztal 2000).

Correlations between sire slopes and GEBV, and accuracies of genomic prediction are presented in Table 1. When only genotyped sires were included in the reference population, accuracies of genomic prediction ranged from 0.38 to 0.53 in Holsteins and 0.49 to 0.63 in Jerseys. When the reference set was expanded to include genotyped cows, the accuracies of genomic predictions showed a slight increase in some cases but not all.

The accuracies of genomic predictions for heat tolerance we have reported indicate that genomic selection offers a promising tool to predict heat tolerance for individual animals based on their genotypes. This will enable farmers to improve the milk production at higher heat load conditions of their herd over time through selection decisions.

ACKNOWLEDGEMENTS

The authors thank the Department of Agriculture of Australia (Canberra, Australia) and Dairy Futures CRC for funding this project and the Australia Dairy Herd Improvement Scheme (Melbourne, Australia) for providing production data.

REFERENCES

Armstrong D. V. (1994). *J. Dairy Sci.* **77**: 2044.

Bernabucci U., Biffani S., Buggiotti L., Vitali A., Lacetera N. and Nardone A. (2014). *J. Dairy Sci.* **97**: 471.

Browning B. L. and Browning S. R. (2009). *Am. J. Hum. Genet.* **84**: 210.

Garrick D. J., Taylor J. F. and Fernando R. L. (2009). *Genet. Select. Evol.* **41**:1.

Gilmour, A. R., Gogel B. J., Cullis B. R. and Thompson R. (2009). VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Haile-Mariam, M., Carrick M. J. and Goddard M. E. (2008) *J. Dairy Sci.* **91**: 4840.

Hayes, B. J., Bowman P. J., Chamberlain A. J., Savin K., van Tassell C. P., Sonstegard T. S. and Goddard M. E. (2009) *PLOS One* **4**: 1.

Hayes, B. J., Carrick M., Bowman P. and Goddard M. E. (2003) *J. Dairy Sci.* **86**: 3736.

Ravagnolo, O. and Misztal I. (2000) *J Dairy Sci* **83**: 2126.

St-Pierre, N. R., Cobanov B. and Schnitkey G. (2003) *J. Dairy Sci.* **86**: 52.

Yang, J., Benyamin B., McEvoy B. P., Gordon S., Henders A. K., Nyholt D. R., Madden P. A., Heath A. C., Martin N. G., Montgomery G. W., Goddard M. E. and Visscher P. M. (2010) *Nat. Genet.* **42**:565.

IMPROVING THE ACCURACY OF GENOMIC SELECTION FOR LACTATION ANOESTROUS INTERVAL IN BRAHMAN CATTLE

M.L. Wolcott, Y.D. Zhang and D.J. Johnston

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351, Australia

SUMMARY

Lactation anoestrous interval (time from the beginning of the annual mating period to the identification of the cows' first subsequent *corpus luteum* by ultrasound scanning) was measured in 898 Brahman females. Previous analyses, including genomic prediction, were performed using a trait defined in cows which successfully weaned a calf (as 3 year olds) from their first mating (LAI: n = 629). The current study expanded this dataset by including an additional 269 records for cows from the same experiment, whose first calf was weaned from their second annual mating, increasing the number of records for the new trait (LAI12) to 898. Heritability for LAI12 (0.43 ± 0.13) was consistent with the previous estimate for LAI (0.51 ± 0.18). A genome wide association study identified more significant SNPs at the $P < 0.01$ level for LAI12 (n = 16,886) than were previously identified for LAI (n = 597). Importantly, a five-fold cross-validation analysis showed that the accuracy of genomic EBVs was increased from 0.14 to 0.24. Expanding the definition of lactation anoestrous interval to that described for LAI12 identified more significant SNPs associated with the trait and increased the accuracy of the associated genomic EBVs for lactation anoestrous interval.

INTRODUCTION

Female reproduction is a key driver of profitability for beef producers in northern Australia. Research examining the genetics of reproductive traits in northern Australia's tropically adapted beef cattle has shown that extended post-partum anoestrous intervals contributed significantly to low weaning rates (Baker 1969; Entwistle 1983). This was more recently confirmed by Johnston *et al*. (2014) who showed that lactation anoestrous (LAI: a measure in lactating first calf females of the time in days between the start of the second annual mating period and the estimated date of first ovulation based on regular ultrasound scanning to detect ovarian function) was more heritable ($h^2 = 0.51$) than measures of net female reproduction traits. The study also showed that in Brahman females, lower (more favourable) LAI was significantly genetically associated with higher lifetime calving and weaning rates ($r_g = -0.71 \pm 0.21$ and -0.62 ± 0.24 respectively) in Brahman cows. However, in the cows examined for that study, reproductive rates from their first mating meant that only 63% of females (n = 629) were eligible to receive a LAI record.

Hawken *et al*. (2012) conducted a genome wide association study to identify single nucleotide polymorphisms (SNPs) which explained genetic variation for a range of female reproduction traits in the same females described by Johnston *et al*. (2014), and Zhang *et al*. (2014) estimated genomic breeding values, and their accuracy, from the same data. This research showed that some opportunities exist to exploit genomic selection for female reproduction, but that LAI had the lowest accuracy of genomic prediction (0.14) due, in part at least, to low numbers of records for the trait. This study, therefore, aimed to maximise the data which could be analysed for lactation anoestrous interval in Brahman females by re-defining a trait to include records from cows which failed to wean a calf from their first annual mating, to estimate variance components for this trait and to determine whether it can improve the accuracy of genomic prediction in Brahman cattle.

---

[*] AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

MATERIALS AND METHODS

Animals and measurements. Animals evaluated for this study were a subset of the Brahman females (N = 1,026) from the Co-operative Research Centre for Beef Genetic Technologies Northern Project. Breeding and management of heifers up to their first annual mating was described by Barwick *et al*. (2009), and Johnston *et al*. (2009; 2014) described management through their first (as 2 year olds), and subsequent matings. Cows which failed to wean a calf in consecutive years were removed from the experiment, which meant that no culling on the basis of reproductive performance was undertaken in females prior to their second annual mating. At the start of the second and subsequent annual mating period, regular ultrasound scanning of the ovaries to identify the presence of a *corpus luteum* was undertaken in lactating cows. This identified the onset of cycling and this interval defined LAI (Johnston *et al*. 2014). For the current analysis, females whose first calving and lactation was the result of their mating as 3 year olds (i.e. for cows which failed to wean a calf from their first mating) also had a record included in the analysis of lactation anoestrous interval. All records for this trait (LAI12) were for the animal's first lactation, and no cow had more than one record included in the analysis.

Statistical analysis. Fixed effect modelling for LAI12 followed the protocols described by Johnston *et al*. (2014) for LAI, and initially included descriptors of the cow (property of origin (4 levels), month of birth (5 levels), dam age (7 levels), mating group (26 levels)), and the calf (calf month of birth (5 levels) calf sex (2 levels)), and all first order interactions. For females whose LAI12 record was from their mating as 3 year olds, mating group described both current and previous mating groups, ensuring that cows which had a record from their first mating were never analysed in the same contemporary group as those whose record came from their second mating. Terms were sequentially dropped from the model based in order of non-significance, with final models containing effects which were significant descriptors of systematic variation in LAI12 at $P < 0.05$. Following the methods described by Johnston *et al*. (2014), variance components for LAI and LAI12 were calculated in ASReml (Gilmour *et al*. 2009), with animal fitted as random and relationships between animals described using a three generation pedigree.

Genotyping and genome wide association study (GWAS). The Beef CRC database includes high density Illumina genotypes (HD: 729,068 SNPs) for 1137 animals, with a further 14, 110 imputed to this level from the results of Illumina 50K SNP chips using the BEAGLE software package, with an accuracy of 90% (as described by Bolormaa *et al*. 2013). Of the 898 BRAH females with a record for LAI12, 875 had genotypes which could be analysed for this study. Thirteen percent of these were from the HD SNP chip, with the remainder imputed from 50K. SNPs with low minor allele frequencies (< 0.05) were excluded from the analyses. The magnitude and significance of SNP effects were estimated as the solutions for each SNP when fitted as a fixed effect in a model with animal fitted as random and including all significant descriptors of environmental variation. The expected false discovery rate (FDR) was calculated as $FDR = p(1-s/t)/[(s/t)(1-p)]$, where p is the threshold significance level tested (e.g. 0.001), s is the number of significant markers, and t is the total number of markers evaluated.

Genomic estimated breeding values (GEBV) for LAI12 were calculated using genomic best linear unbiased prediction (GBLUP). The genomic relationship matrix was fitted as a random effect to estimate GEBVs (Zhang *et al*. 2014), using the method described by Yang *et al.* (2010), and inverted using the Wombat software package (Meyer 2007). GEBV Accuracy (ACC) was estimated as $ACC = r / h$, where r is the correlation between GEBVs and phenotypes and h is the square root of the heritability. The accuracy reported for this study is the mean of five estimates from a five-fold cross validation of GEBV estimates (Zhang *et al*. 2014).

RESULTS AND DISCUSSION

Table 1 presents the number of observations and descriptive statistics for LAI (from Johnston *et al*. (2014)) and LAI12. By including results for Brahman cows whose first lactation was from their second annual mating, the number of records available for analysis increased from 629 to 898 (by 43%), with LAI12 displaying a lower mean and slightly lower standard deviation than LAI. This shows that when Brahman cows' first lactation was from their second annual mating, lactation anoestrous interval was shorter than for cows when the trait was from their first annual mating. As more data becomes available for lactation anoestrous, it may be appropriate to test whether first lactation anoestrus records from the first and second annual mating were the same trait, and consider running a bivariate analyses if the relationship proved to be less than unity.

Table 1. Number of records (N), mean and standard deviation (s.d.), with additive and phenotypic variances (σ^2_a and σ^2_p) and heritability (h^2) (standard error in parenthesis) for LAI and LAI12.

Trait	N	Mean	s.d.	σ^2_a	σ^2_p	h^2
LAI† (days)	629	134	109	5238	10271	0.51 (0.18)
LAI12 (days)	898	116	106	4115	9482	0.43 (0.13)

† From Johnston *et al*. (2014)

Table 1 also presents the additive and phenotypic variances for LAI and LAI12, and the resultant heritabilities for the traits. These show that expanding the definition of lactation anoestrous did not significantly change its heritability suggesting that selection to improve the trait would be similarly effective for LAI and LAI12.

Genome wide assessment of SNP effects for LAI12. The numbers of significant SNPs at levels from $P < 0.01$ to 0.00001, and the proportion in each of these categories expected to be identified by chance, are presented in Table 2. The magnitudes of SNP significance estimated in the current experiment were, higher than those reported for LAI in the study of Hawken *et al*. (2012), where only 530, 66 and 3 SNPs significant at the $p < 0.01$, 0.001 and 0.0001 levels were reported.

Table 2. Number of significant SNPs from genome wide association study (GWAS) for LAI12 in Brahman cows, and the proportion which could be expected as false positives (FDR) at significance levels from $P < 0.01$ to $P < 0.00001$.

GWAS outcomes	N (MAF > 0.05) ‡	$P < 0.01$	$P < 0.001$	$P < 0.0001$	$P < 0.00001$
SNPs Observed	567,445	14,449	2,127	289	21
FDR		0.38	0.27	0.20	0.27

‡ Total number of SNPs included in the analysis with minor allele frequency greater than 0.05

Similarly, the FDR for LAI12 from the current study represents a marked improvement for those reported by Hawken *et al*. (2012) for LAI at the 0.001 and 0.0001 levels of 0.67 and 1.00 respectively. These results suggest that expanding the definition of lactation anoestrous to include results for cows whose first calf was from their second annual mating, increased the capacity of genotypic data to describe genetic variation in lactation anoestrous in Brahman females.

Accuracy of genomic estimated breeding values for LAI12. The results of the five-fold cross-validation showed that the accuracies for GEBV for LAI12 in Brahman was 0.24. This result

was higher than the accuracy of 0.14 reported by Zhang *et al*. (2014) when the trait was defined exclusively in cows which had successfully weaned a calf from their first annual mating. Given the high inputs in management, operator expertise and therefore cost associated with accurate measurement of lactation anoestrous, the opportunity to evaluate the trait using genomic technologies is particularly attractive. The results of this study show that by expanding the definition of the trait to include data from cows whose first lactation was from their second annual mating, the accuracy of genomic breeding values was increased by a factor of almost 60%.

CONCLUSIONS

Lactation anoestrous is an important determinant of reproductive performance in northern Australia's Brahman cattle population. Results of this study have shown that expanding the definition of the trait to include results from cows whose first lactation was from their second annual mating, increased the number of records available for analysis while not significantly changing heritability for the trait. A genome wide association study showed that more significant SNPs were identified for LAI12 than were found for LAI. Importantly, the accuracy of genomic breeding values estimated for LAI12 were also significantly higher than those estimated for LAI. Lactation anoestrous is a difficult and time consuming trait to measure accurately, particularly under the extensive conditions which prevail in northern Australia. As more data becomes available for the trait from research, beef information nucleus and seedstock herds, genomics will provide opportunities to improve lactation anoestrous in Brahman cattle by selection.

ACKNOWLEDGEMENTS

The authors acknowledge Meat and Livestock Australia for their financial support through B.NBP.0761, and the CRC for Beef Genetic Technologies which generated the data analysed for this study.

REFERENCES

Baker A.A. (1969) *Aust. Vet. J.* **45**: 180.

Barwick S.A., Johnston D.J., Burrow H.M., Holroyd R.G., Fordyce G., *et al*. (2009) *Anim. Prod. Sci.* **49**: 367.

Bolormaa S., Pryce J.E., Kemper K., Savin K., Hayes B.J., *et al*. (2013) *J. Anim. Sci.* **91**: 3088.

Entwistle K.W. (1983) *Australian Meat Research Committee Review* **43**: 1.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) 'ASReml user guide. Release 3.0.' VSN International: Hemel Hempstead, UK.

Hawken R.J., Zhang Y.D., Fortes M.R.S., Collis E., Barris W.C., *et al*. (2012) *J. Anim. Sci.* **90**: 1398.

Johnston D.J., Barwick S.A., Fordyce G., Holroyd R.G., Williams P.J., *et al*. (2014) *Anim. Prod. Sci.* **54**: 1.

Johnston D.J., Barwick S.A., Corbet N.J., Fordyce G., Holroyd R.G., *et al*. (2009) *Anim. Prod. Sci.* **49**: 399.

Meyer K. (2007) *Journal of Zhejiang University, SCIENCE B.* **8**: 815.

Yang J., Benyamin B., McEvoy N.P., Gordon S., Henders A.K., *et al*. (2010) *Nature Genetics* **42**: 565.

Zhang Y.D, Johnston D.J., Bolormaa S., Hawken R.J. and Tier B. (2014) *Anim. Prod. Sci.* **54**: 16.

AUSTRALIAN SHEEP BREEDING VALUES FOR WORM EGG COUNT RETAIN PREDICTIVE POWER ACROSS FLOCKS IN THE PRESENCE OF GxE

L. Li[1,3], A.A. Swan[1,3], D.J. Brown[1,3] and J.H.J. van der Werf[2,3]

[1]Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351
[2]School of Environmental and Rural Science, University of New England, Armidale, NSW, 2351
[3]Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW, 2351

SUMMARY

Genotype by environment interactions (GxE) for worm egg count (WEC) in Merino sheep were estimated in eight environments across Australia from the Sheep CRC Information Nucleus flock (IN). Genetic correlations between environments were estimated using a factor analytic model, with mean correlations for each environment ranging from 0.27 to 0.57 for an overall mean of 0.40, confirming the presence of large GxE effects for WEC. The industry genetic evaluation model for WEC fits a direct genetic effect averaged across environments, which is reported back to breeders as the Australian Sheep Breeding Value (ASBV), with a sire by environment interaction term to accommodate deviations in performance (not reported to breeders). This model was validated using the IN data, with results demonstrating that the average genetic effect does retain predictive power across environments, albeit with lower accuracy due to a lower heritability observed in the sire interaction model when GxE effects are large.

INTRODUCTION

Gastrointestinal parasites cause significant economic losses to the Australian sheep industry, and part of the integrated control strategy to reduce these losses is selection of sheep which are resistant to infection (e.g. Eady *et al.* 1996; Gray 1997). The MERINOSELECT and LAMBPLAN across flock genetic evaluation services (Brown *et al.* 2007) provide Australian Sheep Breeding Values (ASBVs) for worm egg count (WEC), and these allow ram breeders to identify genetically resistant sheep. WEC is a highly variable trait and measurements in different environments are affected by a number of different factors, including climatic conditions, worm species, treatment strategies, grazing management and host-parasite interactions. Previous studies based on the MERINOSELECT database have shown significant GxE (Pollot and Greef 2004; Carrick and van der Werf 2007), but are limited by the number of sires used across environments. The Sheep CRC Information Nucleus (van der Werf *et al.* 2010) is an ideal resource to study GxE with a large number of sires progeny tested across eight locations that represent the diversity of Australian sheep production environments. In this study we estimated genetic correlations for WEC across the eight "environments" in the Information Nucleus, and evaluate the impact of significant GxE on the genetic evaluation model used to estimate WEC ASBVs.

MATERIALS AND METHODS

Information Nucleus data description. Worm eggs were counted using a modified McMaster technique and included three species, *H. contortus*, *T. colubriformis*, and *T. circumcincta*. Faecal samples were collected from individual animals when the average of their cohort group exceeded a threshold of 500 eggs per gram (epg). The analyses included 8,509 records from the post-weaning stage (average age 131 days, with a range of 61 to 222 days), collected between 2007 and 2012. The animals represented were the progeny of Merino, Dohne Merino, and SAMM sires mated to Merino dams. They were located at eight sites across Australia and these represent the diversity of

* AGBU is a joint venture of NSW Dept. of Primary Industry and the University of New England

sheep production environments, and for the purposes of this study sites are defined as environments in the GxE sense. A summary of the numbers of animals and sires represented at each site is shown in Table 1. The total number of sires in the study was 308 and the number of sires used across pairs of sites ranged from 24 to 184. Where available, larval species differentiation by site and year of birth showed that a mixture of *T. colubriformis* and *T. circumcincta* was most common, with *H. contortus* observed in significant numbers only at two sites in a single year for each. WEC data were transferred to the cube root scale for analysis.

Estimation of the genetic correlation between environments. Factor analytic models which are known for the parsimonious description of covariance structures and computational advantages (Meyer 2009) were fitted to the data from all sites using the ASReml software package (Gilmour *et al.* 2009). Fixed effects included contemporary group, formed using management group, site, year, sex, breed type (Merino, Dohne, SAMM) and date of measurement (252 levels), birth type (5 levels: 1-5), rearing type (3 levels: 1-3), age of measurement (in days) fitted as a covariate and dam age (2-10 years) fitted as linear and quadratic covariates. The random sire × site effect was modelled with a factor analytic covariance structure (FA). A model with a single common factor was selected on the basis of the log-likelihood ratio tests. Heterogeneous residual variance was fitted in the model at the site level.

Evaluating the impact of GxE in across-flock genetic evaluations. The MERINOSELECT and LAMBPLAN genetic evaluation systems analyse WEC in a multi-trait model, where the traits are defined by age of measurement in four age stages: weaning, post-weaning, yearling and hogget WEC. For each stage, the model includes a direct additive genetic effect and a sire by flock-year interaction effect. The direct additive genetic effect is reported as the WEC ASBV, while the sire interaction effect is included in the model to account for deviations in performance. These can arise in individual flock-year subclasses due to effects such as GxE, incomplete recording and preferential treatment of sire progeny groups, and is not reported to ram breeders. With this approach, the evaluation model is capable of adjusting for GxE to a degree, such that ASBVs represent an average genetic merit across the environments in which animals are evaluated. Provided the effects of GxE are not too large, it is thought that this is a reasonable approach.

In order to quantify the predictive power retained by WEC estimated breeding values (EBV) in the presence of GxE, we tested this model using a cross-validation procedure using IN post-weaning WEC data. Firstly, we fitted a single trait animal model including a random sire x site interaction term, with fixed effects as described above, to data from seven of the eight sites. We then calculate the regression of progeny performance in the eighth site on the sire EBVs from the seven site analysis, which has an expected value of 0.5. The process was then repeated for all sites.

RESULTS AND DISCUSSION

The results from univariate analyses for each site show that mean $WEC^{0.33}$ was considerably higher for Kirby and Turretfield (>10 $epg^{0.33}$) than the other sites which averaged 7.3 $epg^{0.33}$ (Table 1). Phenotypic variances also differ significantly across sites, ranging from 3.42 for Rutherglen to 8.34 for Struan. Heritability estimates were low to moderate across sites, ranging from 0.05 ± 0.11 for Hamilton to 0.58 ± 0.07 for Katanning. Generally low to moderate genetic correlations were observed across sites, with the mean correlation for each site with all other sites varying from 0.27 to 0.57 (Table 1), and correlations between individual pairs of sites ranging from 0.21 to 0.85 (not shown). Although the standard errors of genetic correlation estimates were comparatively high due to the relatively small dataset, the mean estimates for each site with all other sites shown in the last column of Table 1, were significantly lower than 0.8 for four out of eight sites. With 0.8 the commonly accepted threshold considered to show biological importance (Robertson, 1959), the results confirm the presence of significant GxE for WEC. This is consistent with two other studies on GxE in Merinos (Pollot and Greef 2004; Carrick and van der Werf 2007). One of the alternative

methods to account for GxE in the genetic evaluation system is the multiple-trait MACE method (Schaeffer, 1994). However, the main difficulty for this trait is how to classify environments, given that there is no obvious pattern in this dataset in terms of the associations between the extent of genetic correlation and geographical or climatic information (results not shown).

Table 1. Number of records, sires, mean of $WEC^{0.33}$, and estimates of the phenotypic variance (V_p), heritability (h^2), and mean GxE genetic correlation (r_g) between environments at each site. Standard errors in brackets (s.e.)

Site	Location	Records	Sires	Mean	V_p (s.e.)	h^2 (s.e.)	Mean r_g (s.e.)[A]
1	Kirby NSW	2482	296	10.8	8.16 (0.24)	0.22 (0.05)	**0.36** (0.19)
2	Trangie NSW	672	69	6.0	4.97 (0.28)	0.21 (0.08)	0.46 (0.25)
3	Cowra NSW	575	89	6.9	6.20 (0.38)	0.21 (0.10)	0.49 (0.26)
4	Rutherglen VIC	845	106	7.7	3.42 (0.18)	0.26 (0.08)	0.39 (0.21)
5	Hamilton VIC	507	71	8.3	4.49 (0.29)	0.05 (0.11)	0.57 (0.23)
6	Struan SA	721	103	6.6	8.34 (0.47)	0.27 (0.11)	**0.31** (0.22)
7	Turretfield SA	1066	110	10.6	5.38 (0.25)	0.31 (0.07)	**0.31** (0.18)
8	Katanning WA	1641	196	8.0	5.85 (0.23)	0.58 (0.07)	**0.27** (0.14)
Mean		1064	130	8.1	5.85	0.26	0.40

[A]Genetic correlation estimates in bold are significantly less than 0.8 at p=0.05 level.

Cross-validation results are shown in Table 2, and on average the regression of offspring performance on sire EBVs calculated in other environments was exactly 0.5, although there was a large range (0.33 to 0.81). This demonstrates that EBVs from a genetic evaluation model fitting sire interaction effects do have predictive power across environments in the presence of significant GxE. We note however that when compared to the average within-environment heritability estimate (0.26 in Table 1), heritability estimates were significantly lower from the single trait sire interaction model fitted across sites (0.09 in Table 2). This can be interpreted by extension of the co-heritability concept from the theory of correlated response (e.g. Falconer and Mackay, 1996): the co-heritability for selection in environment X targeting response in environment Y can be viewed as $\boldsymbol{h_X h_Y r_g}$ where $\boldsymbol{h_X}$ and $\boldsymbol{h_Y}$ are the square roots of heritability in each environment and $\boldsymbol{r_g}$ is the GxE genetic correlation between environments. With an average heritability of 0.26 and average GxE genetic correlation of 0.40 from the results shown in Table 1, the co-heritability has an approximate value of 0.10 in these data. This is very similar to the average heritability estimate shown in Table 2. So while these results show that WEC ASBVs from MERINOSELECT and LAMBPLAN are likely to have predictive power across flocks even in the presence of significant GxE, they will have lower accuracy in an across flock context. For a co-heritability of 0.10 and within environment heritability of 0.26, the reduction in accuracy based on own performance is approximately 38% (calculated from $\sqrt{0.10}/\sqrt{0.26}$). For progeny-tested sires the reduction in accuracy will be lower as the number of progeny increases, especially if these progeny are represented across different environments.

The MERINOSELECT model for post-weaning WEC assumes a heritability of 0.2 and a sire x site interaction variance ratio of 0.02, considerably different to the average estimates of 0.09 and 0.06 of the same parameters in Table 2. It is likely that the difference is due to data structure: in the IN data the majority of sires are used across sites, whereas in the MERINOSELECT data for approximately 60% of the sires that have progeny with WEC measurements, the progeny were recorded in one flock only, and so GxE effects are not represented in a large part of the data.

Table 2. Cross-validation results for each environment, where V_p (phenotypic variance), h^2 (heritability), and s^2 (sire by site interaction variance ratio) are estimated from a single trait analysis of data for all sites *excluding* the site shown in each row, and b is the regression of offspring performance at the site shown in each row on sire EBVs calculated from all other sites. Standard errors in brackets (s.e.)

Site	Location	V_p (s.e.)	h^2 (s.e.)	s^2 (s.e.)	b (s.e.)
1	Kirby NSW	5.70 (0.12)	0.10 (0.03)	0.07 (0.01)	0.81 (0.19)
2	Trangie NSW	6.56 (0.12)	0.09 (0.02)	0.06 (0.01)	0.33 (0.17)
3	Cowra NSW	6.46 (0.12)	0.09 (0.02)	0.06 (0.01)	0.66 (0.21)
4	Rutherglen VIC	6.76 (0.13)	0.08 (0.02)	0.06 (0.01)	0.39 (0.14)
5	Hamilton VIC	6.55 (0.12)	0.07 (0.02)	0.07 (0.01)	0.63 (0.21)
6	Struan SA	6.30 (0.12)	0.10 (0.02)	0.06 (0.01)	0.33 (0.20)
7	Turretfield SA	6.60 (0.13)	0.09 (0.02)	0.06 (0.01)	0.33 (0.13)
8	Katanning WA	6.54 (0.13)	0.08 (0.02)	0.04 (0.01)	0.53 (0.12)
Mean		6.43	0.09	0.06	0.50

The data used in this study have also been used to develop genomic predictions which are used in the calculation of WEC ASBVs. For Merinos the accuracy of these genomic predictions is estimated to be 0.26 (Swan *et al.*, 2014). The genomic analyses are based on IN data from all sites, and it would be worthwhile to investigate the impact of GxE on these analyses.

CONCLUSIONS

The results presented demonstrate that there are significant GxE for WEC in Merino sheep, but that the analysis method used in industry genetic evaluations can account for these to a degree. Although the accuracy of breeding values is most likely to be lower in an across flock context, ram breeders can have confidence in ASBVs based on performance data collected in their own flocks, and for sires with large numbers of progeny tested across a range of environments.

ACKNOWLEDGEMENTS

This work was partially funded by MLA Project B.SGN.0127.

REFERENCES

Brown D.J., Huisman A.E., Swan A.A., Graser H-U., Woolaston R.R., *et al.* (2007) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17:** 187.

Carrick M.J. and van der Werf J.H.J. (2007) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17:** 248.

Gilmour A.R., Gogel B.J., Cullis B.R., and Thompson R. (2009) ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Falconer D.S. and Mackay T.F.C. (1996) 'Introduce to quantitative genetics' 4th ed. Longmans, Essex, UK.

Gray G.D. (1997*) Veterinary Parasitology* **72:** 345.

Pollott G.E. and Greeff J.C. (2004) *J. Anim. Sci.* **82:** 2840.

Robertson A. (1959) *Biometrics* **15:** 469.

Swan A. A., Brown D.J., Daetwyler H. D., Hayes B.J., Kelly M.J., *et al.* (2014) *Proc. 10th World Congr. Genet. Appl. Livest. Prod.*, paper 334.

Meyer, K. (2009) *Genet. Sel. Evol.* **41**: 21.

Van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Ani. Prod. Sci.* **50:** 998.

BESSiE
A PROGRAM FOR MULTIVARIATE LINEAR MODEL BLUP AND BAYESIAN ANALYSIS OF LARGE SCALE GENOMIC DATA.

V. Boerner and B. Tier

Animal Genetics and Breeding Unit[1], University of New England
Armidale, 2351, NSW, Australia

SUMMARY

BESSiE is a software designed for uni- and multivariate analysis of linear mixed models including large scale genomic data.

BESSiE facilitates models allowing for various fixed and random effects, and for observations on continuous or categorical scales, and implements different Bayesian algorithms for the prediction of effects of genetic markers (e.g. BayesA, BayesB, BayesCπ and BayesR), GBLUP and SNP-BLUP.

INTRODUCTION

Various software packages are available for the analysis of phenotypic observations with linear mixed models in quantitative genetics, which can be categorised by the employed algorithm for inferring dispersion and location parameters of the modelled factors: a) Restricted Maximum Likelihood (REML) based software, and b) Bayesian and Markov Chain Monte Carlo (MCMC) based software. While various REML software packages specifically designed for quantitative genetics are widely used and well documented,(e.g. ASREML (Gilmour *et al.* 2009), WOMBAT,(Meyer 2007), DMU (Madsen *et al.* 2014), REMLF90 (Misztal *et al.* 2002), VCE (Groeneveld *et al.* 2010)), software packages employing Bayesian and MCMC methodology are less common (GIBBSF90 and THRGIBBSF90 (Misztal *et al.* 2002),BAYESR (Moser *et al.* 2015), MCMCglmm,(Hadfield 2010)), but only GIBBSF90 AND THRGIBBSF90 are explicitly designed for the application on large data sets and complex models in quantitative genetics, and therefore provide results in a reasonable amount of time. The relatively small number of Bayesian and MCMC software packages for quantitative geneticist may reflect the disadvantage of this methodology in terms of processing time. In addition, large scale genomic marker data emerging in the late 2000 easily fit into existing REML software via approaches like GBLUP or single marker regression. In contrast several Bayesian algorithms for sampling dispersion and location parameters of genomic markers have been proposed (Meuwissen *et al.* 2001; Habier *et al.* 2011; Erbe *et al.* 2012), which differ only slightly but require adjustments in the software source code, thus making it more difficult to develop and maintain a software which covers all.

The aim of this paper is to describe the software BESSiE which is designed for uni- and multivariate BLUP and Bayesian analysis of linear mixed models in quantitative genetics allowing for various factors, algorithms, large scale genomic data and both continuous as well as categorical observations.

SOFTWARE DESCRIPTION

BESSiE is written in FORTRAN90, command line operated, parameter file driven and

[1]A joint venture of the NSW Department of Primary Industry and the University of New England

comes with an extensive manual. It is available for 64bit Unix-like operation systems only, and is optimised for Intel architecture.

The super-set model to be fitted in BESSiE may be written as:

$$\begin{pmatrix} y_1 \\ . \\ y_n \end{pmatrix} = \begin{pmatrix} X_1 & . & 0 \\ . & . & . \\ 0 & . & X_n \end{pmatrix} \begin{pmatrix} b_1 \\ . \\ b_n \end{pmatrix} + \begin{pmatrix} Z_1 & . & 0 \\ . & . & . \\ 0 & . & Z_n \end{pmatrix} \begin{pmatrix} u_1 \\ . \\ u_n \end{pmatrix} + \begin{pmatrix} Q_1 M & . & 0 \\ . & . & . \\ 0 & . & Q_n M \end{pmatrix} \begin{pmatrix} g_1 \\ . \\ g_n \end{pmatrix} + \begin{pmatrix} e_1 \\ . \\ e_n \end{pmatrix}$$

where $(y_1,.,y_n)'$, $(b_1,.,b_n)'$, $(u_1,.,u_n)'$, $(g_1,.,g_n)'$ and $(e_1,.,e_n)'$ are vectors of phenotypic observations of linear or categorical scale, fixed effects, random non-marker effects and random marker effects, X, Z and Q are matrices relating the effects to their respective observations, M is a matrix of marker genotypes of dimension "number of genotyped individuals"×"number of markers" and the subscripts are for trait 1 to n. Values in X may be dummy variables or linear co-variables, where for the latter the order of polynomial regression is user-defined. Values in $(u_1,.,u_n)$ are assumed to be distributed $N([0,.,0]', A\otimes\Sigma)$, $N([0,.,0]', G\otimes\Sigma)$, $N([0,.,0]', I\otimes\Sigma)$ or $N([0,.,0]', K\otimes\Sigma)$, where A is the pedigree derived numerator relationship matrix, G is a relationship matrix derived from genetic markers, I is an identity matrix, K is an unknown matrix of dimension "number of factor levels"× "number of factor levels" provided by the user, and Σ is a co-variance matrix of factors. Note that all random non-marker effects can be fitted together.

The algorithm to obtain dispersion and location parameters when trait observations are of categorical scale is described in Sorensen *et al.* (1995) and Albert and Chib (1993).

Random effects of genetic markers $(g_1,.,g_n)'$ can be obtained from BayesA and BayesB (Meuwissen *et al.* 2001), BayesCπ (Habier *et al.* 2011), BayesR (Erbe *et al.* 2012) or ridge regression SNP-BLUP (Piepho 2009). For BayesA, BayesB and BayesCπ, all relevant parameters of the algorithms and the prior distributions of marker variances are taken from the related publications, but can also be set by the user.

Residuals are assumed to be distributed $N([0,.,0]', I\otimes R)$, where R is the residual co-variance matrix of dimension $n \times n$. However, to account for observations with different residual variances (e.g. de-regressed breeding values), a co-variance Ω can be modelled, where Ω is a block-diagonal matrix containing $\omega_1\sigma^2_{e_1}$ to $\omega_n\sigma^2_{e_n}$ in the diagonal elements of the diagonal blocks, and $\sqrt{\omega_1\omega_n}\sigma_{e_{1,n}}$ in the diagonal elements of the off-diagonal block which links trait 1 and trait n, where ω_1 and ω_n are the weights of trait 1 and n, and $\sigma^2_{e_1}$, $\sigma^2_{e_n}$ and $\sigma_{e_{1,n}}$ are the residual variances and co-variance of both the traits.

In multivariate analysis using BayesA, BayesB, BayesCπ or BayesR effects of genetic markers are estimated from

$$\left(\begin{bmatrix} Q_1 M & . & 0 \\ . & . & . \\ 0 & . & Q_n M \end{bmatrix}' R^{-1} \begin{bmatrix} Q_1 M & . & 0 \\ . & . & . \\ 0 & . & Q_n M \end{bmatrix} + \begin{bmatrix} \sigma^2_1 & . & 0 \\ . & . & . \\ 0 & . & \sigma^2_n \end{bmatrix}^{-1} \right) \begin{pmatrix} g_1 \\ . \\ g_n \end{pmatrix} =$$

$$\begin{pmatrix} Q_1 M & . & 0 \\ . & . & . \\ 0 & . & Q_n M \end{pmatrix}' R^{-1} \left(\begin{bmatrix} y_1 \\ . \\ y_n \end{bmatrix} - \begin{bmatrix} X_1 & . & 0 \\ . & . & . \\ 0 & . & X_n \end{bmatrix} \begin{bmatrix} b_1 \\ . \\ b_n \end{bmatrix} - \begin{bmatrix} Z_1 & . & 0 \\ . & . & . \\ 0 & . & Z_n \end{bmatrix} \begin{bmatrix} u_1 \\ . \\ u_n \end{bmatrix} \right)$$

where σ^2_1 to σ^2_n are diagonal matrices of dimension "number of markers"× "number of markers" of which elements contain the marker variances generated according to the Bayesian method specified for trait 1 to n. The co-variances between the effects of a genetic marker

on trait 1 to n are assumed to be zero.

BESSiE has no hard coded limitations in terms of number of traits, factors, genotypes and markers, and has been tested on very large data sets.

As an example, a bi-variate analysis with 4,420 individuals genotyped for 510,174 single nucleotide polymorphism (SNP), 19,549 individuals in the pedigree, 7 fixed effects and a polygenic random effect per trait, and SNP effects modelled according to BayesR with 4 distributions requires 4.3GB of RAM and about 7 real time seconds on an Intel(R) Core(TM) i7-3770 processor to sample all location and dispersion parameters once.

BESSiE comes without any warranties and can be used by the scientific community free of charge. It can be downloaded from http://turing.une.edu.au/~agbu-admin/BESSiE/.

REFERENCES

Albert J. H. and Chib S. (1993) *J. Am. Stat. Assoc.* **88**:669 .

Erbe M., Hayes B. J., Matukumalli L. K., Goswami S., Bowman P. J., Reich C. M., Mason B. A. and Goddard M. E. (2012) *J. Dairy Sci.* **95**(7):4114.

Gilmour A. R., Gogel B. J., Cullis B. R. and Thompson R. (2009) ASReml User Guide Release 3.0, VSN International Ltd, Hemel Hempstead.

Groeneveld E., Kovac M. and Mielenz N. (2010) VCE user's guide and reference manual version 6.0.

Habier D., Fernando R. L., Kizilkaya K. and Garrick D. J. (2011) *BMC Bioinformatics* **12**:186.

Hadfield J. D. (2010) *J. Stat. Softw.* **33**(2).

Madsen P., Jensen J., Labouriau R., Christensen O. F. and Sahana G. (2014) in Proc. 10th. WCGALP, Vancouver, Canada.

Meuwissen T. H., Hayes B. J. and Goddard M. E. (2001) *Genetics* **157**(4):1819.

Meyer K. (2007) *J. Zhejiang Univ. Sci. B* **8**(11):815.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T. and Lee D. (2002) in Proc. 7th. WCGALP, Montpellier, France.

Moser G., Lee S. H., Hayes B. J., Goddard M. E., Wray N. R. and Visscher P. M. (2015) *PLoS Genet* **11**(4):e1004969.

Piepho H. P. (2009) *Crop Sci.* **49**(4):1165.

Sorensen D. A., Andersen S., Gianola D. and Korsgaard I. (1995) *Genet. Sel. Evol.* **27**:229 .

Proc. Assoc. Advmt. Breed. Genet. 21: 393-396

COLLECTION OF DATA FOR THE GENETIC IMPROVEMENT OF HEALTH TRAITS IN AUSTRALIAN DAIRY CATTLE

M. Abdelsayed[1], P. Douglas[1] and J.E. Pryce[2]

[1]Holstein Australia, Hawthorn East, VIC, 3122
[2] Department of Economic Development, Jobs, Transport and Resources, Government of Victoria Bundoora, VIC, 3083

SUMMARY

There is growing interest in the Australian dairy industry in the genetic improvement of the health of dairy cows. In Australia, there is minimal storage or export of health data from some on-farm software systems into industry databases to assist in the research or reporting of health traits. The Health Data for Healthy Cows (HDHC) project aims to get a better understanding of the extent of health data recording in the Australian industry by collecting health data from the 100 Ginfo (genomic reference) herds. Health data obtained from herd test centres totalled 275,729 records from just 46 out of the 100 herds. The four most recorded groups of health diseases identified were mastitis, reproductive problems, lameness and metabolic disorders. Mastitis had the highest incidence with 20% of cows affected, followed by reproductive problems (12%), lameness and metabolic disorders (5% and 3% respectively). This project has provided an insight into what health information is actually being collected on farm and that there is a source of health data available which can be accessed and potentially used for the genetic improvement of health traits in Australian herds.

INTRODUCTION

Great improvements have been made genetically in milk production in dairy cows over the last 60 years. However unfavourable genetic relationships between milk production and most disease traits, such as mastitis, lameness, reproductive problems and metabolic disorder health traits have become apparent as milk production has increased (Pryce *et al*. 1997; Rauw *et al*. 1998; Koeck *et al*. 2012).

A growing concern for dairy farmers is the improvement of dairy cow health through genetic selection. Healthy cows are more productive, easier to manage, require less intervention, have improved animal welfare and contribute to profitability by reducing production costs. However, in many countries, including Australia, industry collection of data on common health events has been sub-optimal or absent, which means there is no ability to provide breeding values and apply genetic selection for common health disorders. Also, such traits are low in heritability, meaning that although genetic progress is feasible, it will be slower. While many farmers may collect some of this information on farm, there is likely to be variation in the completeness of these data sets. In Australia, there is little storage or export of such information from some on-farm software packages into industry databases for research or reporting purposes.

Before any work can begin on providing Australian farmers with breeding values for common health disorders, it is important to quantify what data is already being collected on farm and in veterinary practices. As a result of this challenge, the health data for healthy cows (HDHC) project has commenced to help improve our understanding of the extent of health data recording in the Australian dairy industry. The HDHC project will use infrastructure through the Dairy CRC in the form of the 100 ‘Ginfo’ (Genomic information) herds to collect all health data that is currently being amassed on farm. The Ginfo data is being used as a genomic reference population for genomic breeding values. One of the advantages of having a genotyped population is that it opens up new opportunities for new breeding values, such as dairy health traits. Therefore the objectives

of the HDHC project are to:

1. Investigate and identify health data sources available within the herds participating in the Ginfo project
2. Assimilate health data into a database in order to summarise health data status
3. Estimate the incidence of common diseases and health occurrences on dairy farms
4. Estimate antibiotic usage on farms
5. Calculate provisional genetic parameters for health traits where incidence is high enough
6. Estimate the accuracy of genomic selection achievable for data identified health traits

For this paper the first three objectives are covered.

MATERIALS AND METHODS

Health data sources. On farm health data were collected from the 100 Ginfo participating herds.

Survey. A survey was used to get an understanding of the health data collection and storage methods that occur on farm. Before participating in the survey, privacy consent and data release authority documents were also provided to the participants to comply with privacy laws.

Obtaining health data. Once the survey, data release authority and privacy consent forms were returned by the farmers, an email was sent to each of the herds respective herd test centres to request all of the data collected for that farm, including the health data interchange format file.

Data analysis. Each of the herds' health data files were merged together to create a master health data file which contained the national cow ID, herd ID, health event, health treatment, date of health treatment, calving date, breed and cow date of birth. Statistical and graphical summaries of the surveys and master health data file were produced using Microsoft office Excel 2013 and the statistical program R version 3.1.1.

Disease incidence calculation. For the calculation of disease incidence the following equations were used:

Number of disease cases (disease occurrence) =

No. of disease cases over a lactation / No. of cows-lactations with disease

Disease occurrence is calculated for cows with a health event

Number of cows with cases (%) =

No. of cows with cases /Total No. of cows with test records

Cases refers to specific diseases, for this paper cases refers to mastitis, reproductive disorders, lameness and metabolic diseases.

RESULTS AND DISCUSSION

To date 51 of the Ginfo participants have returned surveys. Out of these, 46 herds have health data recorded, while 5 had no health records. Therefore, the results currently presented include 46 of the herds out of the total 100 participating in the HDHC project. The total number of raw health records (before any quality control) obtained only from herd test centres is 275,729 records between 1998 to 2011, from 42, 056 cows, representing multiple treatments per cow.

Survey. The main form of recording of health data on farm is electronically. Health events are mostly recorded daily across the Ginfo herds.

Figure 1 illustrates the health diseases that dairy farmers regard as most important and what they stated that they recorded on farm.

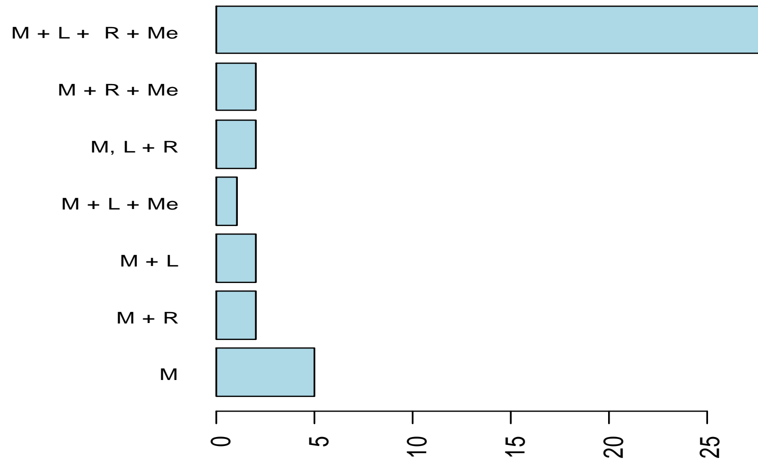


Figure 1. Health diseases farmers "say" they record and which they also regard as important. Mastitis (M), Reproductive (R), Lameness (L) and Metabolic (Me).

Data analysis. Analysis of the data indicated that the general health events most commonly recorded are mastitis, reproductive problems (retained foetal membranes, uterine infections, cystic ovarian disease), lameness (foot and leg disorders, injuries, footrot, abscess) and metabolic disorders (milk fever, ketosis, Grass tetany) (Figure 2). This is fairly consistent with the survey conclusions on what farmers indicated they record and what they think are most important (Figure 1).

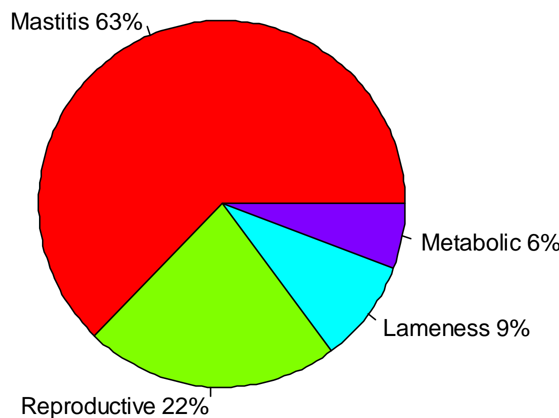


Figure 2. Health data that is actually being recorded on farm.

More mastitis health event data being identified and made available for analysis purposes will assist in improving the reliability of the mastitis resistance breeding value. With fertility being one of the most significant issues facing the dairy industry, knowing that reproduction disorder health data incidences are actually being recorded indicates that farmers are seeing the value and importance of such a trait to the health and production of their cows. Therefore, having a large health event data set would potentially increase the ability for farmers to improve their fertility, and in return profit, through the incorporation into a multi-trait genetic analysis model to improve

the reliability and confidence of the fertility breeding value.

Identifying what health event data is actually recorded on farm opens up new avenues of genetic analysis for potential new traits such as lameness and the development of new individual or integrated breeding values to become available to the industry.

Disease incidence. Disease incidence for the top four most recorded health events; mastitis, reproductive problems lameness and metabolic disorders were calculated (Table 1).

Table 1. Disease incidence of the most recorded health traits, total number of cow's with each disease and the total number of cases for each health trait

Disease	No. of cows	Total no. of cases	Disease occurrence	Cows with cases (%)
Mastitis	8495	21611	2.54	20%
Reproductive	4972	7730	1.55	12%
Lameness	2237	3124	1.40	5%
Metabolic	1425	1951	1.37	3%

For cows recorded with mastitis, on average there were 2.5 cases per lactation and 20% of cows affected. Cows reported with reproductive problems had 1.5 cases per lactation, affecting 12% of cows. At a lower level, lameness and metabolic problems in nominated health event cows have about 1 case per lactation with 5% and 3% of health event recorded cows being affected respectively. Incidences from this data set are less than those previously reported in other studies (Clarkson *et al.* 1996; Espejo *et al.* 2006; Clarkson *et al.* 1996; Parker-Gaddis *et al.* 2012). The number of cow cases for lameness, reproductive and metabolic problems were lower than previously reported while mastitis cow cases were fairly similar to findings reported in Norway (Osteras *et al.* 2007). No conclusions at this point can be drawn about whether the herds used in this study are generally healthier than other herds as only half of the herds were used in this analysis.

CONCLUSIONS

In Australian herds, mastitis, reproductive disorders, lameness and metabolic diseases are the most recorded health events. Mastitis is the most common occurring disease in dairy herds, followed by reproductive disorders, lameness and metabolic problems occur at lower incidence. As a result these findings provide information to make clearer decisions on future research priorities, and contribute a reference data set that may be applied for genomic correlation purposes.

REFERENCES

Clarkson M.J., Downham D.Y., Faull W.B., Hughes J.W., Manson F.J. et al. (1996) *Vet. Rec.* **138**: 563.

Espejo L.A., Endres M.I. and Salfer J.A. (2006) *J. Dairy Sci.* **89**: 3052.

Osteras O., Solbu H., Refsdal A.O., Roalkvam T., Filseth O. and Minsaas A. (2007) *J. Dairy Sci.* **90**: 4483.

Koeck, A., Miglior, F., Kelton, D. F., and Schenkel, F. S. (2012) *J. Dairy Sci*. **95**: 4099.

Parker-Gaddis K.L., J. B. Cole J.B., Clay J.S., and Maltecca C.(2012) *J. Dairy Sci.* **95**:5422

Pryce, J.E., Veerkamp, R.F. Thompson, R., Hill, W.G. and Simm, G. (1997) *Anim. Sci*. **65**: 353.

Rauw, W. M., E. Kanis, E. N. Noordhuizen-Stassen, and F. J. Grommers. (1998) *Livest. Prod. Sci.* **56**:15.

Proc. Assoc. Advmt. Breed. Genet. 21: 397-400

QUANTIFYING BETWEEN ANIMAL VARIATION FOR MALE SEMEN TRAITS AND FEMALE EGG TRAITS IN AN OSTRICH FLOCK USED TO DEVELOP AN ARTIFICIAL INSEMINATION PROTOCOL

S.W.P Cloete[1,2], M. Bonato[2], A.M.J. Smith[2] and I.A. Malecki[2,3]

[1]Western Cape Department of Agriculture, Research & Technology Development Services, Directorate: Animal Sciences, Private bag X1, Elsenburg 7607, South Africa; [2]Department of Animal Sciences, University of Stellenbosch, Stellenbosch 7600, South Africa; [3]School of Animal Biology M085, Faculty of Natural and Agricultural Sciences, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

SUMMARY

The development of a viable protocol for artificial insemination (AI) of ostriches may assist to overcome challenges to a structured breeding programme in this species. Data were obtained from up to 16 males trained to ejaculate in an artificial cloaca and 36 females producing eggs in the absence of males being used as a resource flock to develop an AI protocol. Repeatability estimates across years (termed as permanent environment or pe^2) or within years (termed as temporary environment or te^2) were estimated to assess whether current flock gains could be achieved by selecting particular animals. Male semen volume and libido were mostly affected by long-term pe^2 effects at respectively 0.38 and 0.32, while sperm morphology and sperm concentration mostly depended on te^2 effects, ranging from 0.09 for sperm concentration to 0.41 for the percentage normal spermatozoa. Permanent environmental effects were more important for semen motility traits, ranging from 0.03 for amplitude of the lateral head to 0.20 for straightness. However, most derived estimates did not differ from zero. Analysis of monthly female egg production and average egg weight records yielded estimates of 0.21 and 0.45 for pe^2 and 0.11 and 0.15 for te^2, respectively. Selection for better adaptation to the proposed AI protocol may result in current flock gains for some male and female traits.

INTRODUCTION

Genetic evaluation of ostriches is poorly developed when compared to other farmed livestock species (Cloete *et al*. 2008). Challenges associated with ostriches are a communal nesting system not conducive to the recording of parentage in colony mating flocks, confounded random effects and mate incompatibility when mated as pairs to record pedigrees, a very narrow male:female ratio and the absence of a formal recording and evaluation scheme (Bunter and Cloete 2004; Cloete *et al*. 2008). These challenges could all be alleviated by a workable protocol of artificial insemination (AI), which is studied as a viable alternative to natural mating (Malecki *et al*. 2008; Malecki and Rybnik-Trzaskowska 2011). Such technology also has the potential to reduce the male to female ratio and may add additionally to worker occupational health and safety by reducing the number of dangerous males, as well as bird welfare, by reducing incompatibility between animals paired off (Cloete and Malecki 2011). AI contributed markedly to livestock genetic improvement (Verma *et al*., 2012). It is foreseen that the ostrich industry is likely to benefit from AI in the same way.

Since stress-free collection of semen in ostrich males has been developed (Rybnik *et al*. 2007), there have been a number of distinct advances towards making AI a reality in this species. It has, for instance, been demonstrated that semen can be collected from males trained to the dummy female method twice daily, without compromising ejaculate output and sperm quality (Bonato *et al*. 2011). Also, semen could be collected from trained males all year round, although some seasonal variation do occur (Bonato *et al*. 2014b). Stress-free artificial insemination of trained

ostrich females is possible by using the voluntary crouch principle (Malecki *et al*. 2008; Bonato *et al*. 2014c). Furthermore, inseminated females were demonstrated to produce fertile eggs (Bonato *et al*. 2014c).

However, individual variation is observed for all traits that are likely to be of importance and value in an AI program. The question remains whether between-animal variation can be exploited to select individuals well-adapted to AI. To date, all genetic parameters for ostriches have been derived from a single resource flock, employing a pair-breeding system (Cloete *et al*. 2008). It is unknown whether between-animal variance ratios estimated for pair-bred ostriches would also be applicable to usage in an AI program.

For this reason, in this study, the repeatability of sperm traits in males and egg traits in females was studied in a resource flock maintained to develop and refine a viable AI protocol in this species. Repeatability is the upper limit for heritability in cases where no animal permanent environmental effects are present.

MATERIAL AND METHODS

Location, animals and data collected. The animal resource was maintained at the Oudtshoorn Research Farm of the Western Cape Department of Agriculture near Oudtshoorn in the Klein Karoo region of South Africa. Data were sourced from seven males trained to ejaculate in an artificial cloaca and used in the seasonal variation trial of Bonato et al. (2014b) for the assessment of ejaculate volume, sperm concentration, sperm morphology (the percentage of live, normal and dead spermatozoa) as well as male libido (4-point scale) over a period of 24 months. Sixteen males provided 1169 recorded ejaculates for use in this analysis. A total of 257 semen samples from 10 males collected from 2013 to 2015 were used to assess motility traits by using computer assisted sperm analysis (CASA) software (Microptic S.L. System Version 5.2, Barcelona, Spain). Traits considered were progressively motile and motile spermatozoa; curvilinear, straight line and average path velocity; linearity, straightness and wobble as well as beat cross frequency. These sperm motility traits are commonly related to fertilization success in other species (Suarez and Pacey 2006). Monthly egg production and egg weight (expressed as trait of the female) records were collected over 5 consecutive breeding seasons (from May to December, 2009-2013) in a female flock producing eggs in the absence of males (Bonato *et al*. 2014a). In total, 664 female-month records of 36 females were available for analysis.

Statistical analyses. Repeated records produced by the same male or female were accommodated in two ways by using ASReml software (Gilmour *et al*. 2006): firstly, the variance component of unique animals across years was considered as an indication of the animal permanent environment (pe^2) and secondly, the correspondence of records of unique animals within production years were considered as an indication of animal temporary environment (te^2). The former parameter can be considered to reflect long-term effects of specific animals on traits while te^2 reflect short-term effects. No attempt was made to partition the pe^2 variance component in genetic and permanent environmental effects, given the relatively small size of all databases (<1200 records) and the number of animals recorded (<40). Log likelihood tests were used to assess the significance of random effects added to the fixed effect analytical model. Fixed effects included were year (specific for the respective data sets), month or season and whether samples used for CASA were diluted or not. Fixed effects solutions are not relevant for this paper and therefore not reported.

RESULTS AND DISCUSSION

There is little knowledge about the repeatability of sperm traits in ostrich males under repeat sampling. Between-male differences for sperm traits reported by Bonato *et al*. (2011; 2014b) suggest that such traits are likely to be repeatable. Estimates of pe^2 and te^2 for sperm traits and

libido are presented in Table 1. The results in Table 1 show that pe² effects predominated in semen volume and libido scores. In contrast, te² effects were more important for sperm concentration and morphology. Nonetheless, male specific semen traits were repeatable, suggesting that these records can be used to select males with a high semen yield and/or libido for usage in the current flock. No comparable work on ostriches, other ratites or avian species could be found.

Table 1. Repeatability estimates (±s.e.) for semen volume, sperm concentration, sperm morphology traits and libido for the unique male across years (pe²) and the unique male within years (te²) effect, for males trained for the routine collection of semen using the dummy female method

Trait	pe²	te²
Semen volume (mm)	0.38±0.12	0.13±0.06
Sperm concentration (x10^9/ml)	–	0.09±0.03
Normal sperm (%)	–	0.41±0.08
Abnormal sperm (%)	–	0.31±0.06
Dead sperm (%)	–	0.38±0.07
Libido score (n)	0.32±0.10	–

The between-male variance of male sperm motility traits mostly partitioned towards pe² (Table 2). Derived estimates were, however, relatively low (<0.20), and only reached the level of double the corresponding s.e. for straightness. Although mostly not significant ($P<0.05$) when related to the corresponding s.e., the inclusion of the additional random term improved the log likelihood ratio for all traits except the amplitude of lateral head, and wobble. The pe² variance ratios for all traits associated with velocity were below 0.10. The only trait primarily affected by te² was linearity. We did not find any literature on similar traits in ostriches or other ratites.

Table 2. Repeatability estimates (±s.e.) for male sperm motility traits according to CASA for the unique male across years (pe²) and unique male within years (te²) effect, for males trained for the routine collection of semen using the dummy female method

Trait	pe²	te²
Progressive motile spermatozoa (%)	0.13±0.08	–
Motile spermatozoa (%)	0.19±0.10	–
Curvilinear velocity (µm/s)	0.08±0.07	–
Straight line velocity (µm/s)	0.09±0.07	–
Average path velocity (µm/s)	0.07±0.06	–
Amplitude of lateral head (µm)	0.03±0.05	–
Linearity (%)	–	0.12±0.09
Straightness (%)	0.20±0.10	–
Wobble (%)	0.05±0.05	–
Beat cross frequency (Hz)	0.10±0.07	–

Between-female variance components for egg production and average egg weight partitioned mostly to pe² and to a lesser extent to te² (Table 3). The estimate of pe² were approximately double that of te² for egg production and about three times as high for average egg weight. The derived parameters were close to a pe² estimate of 0.17 (the sum of h², pe² and the service sire variance ratio) and a te² estimate of 0.15 derived by Fair *et al.* (2011) for monthly egg production records of pair-mated ostriches. Corresponding estimates by Fair *et al.* (2011) for egg weight amounted to 0.57 for pe² and 0.13 for te², which also corresponded fairly well with the present estimates. From these results it seems as if parameters derived for a flock of females producing eggs in the absence of males are quite similar to those derived for pair-mated females.

Table 3. Repeatability estimates (±s.e.) for monthly female egg production for the unique female across years (pe²) and unique female within years (te²) effects, for females maintained without males and producing eggs in the absence of males

Trait	pe²	te²
Egg production (n)	0.21±0.07	0.11±0.04
Average egg weight (g)	0.47±0.08	0.15±0.04

CONCLUSION

This study is the first report of between-animal variation available for exploitation in an AI program for ostriches, aimed at overcoming typical production challenges, while also promoting genetic progress as in other species. Key male traits like semen volume and libido were repeatable across years, while sperm morphology traits were more affected by short-term animal affects. Between-animal effects for sperm motility traits were smaller and mostly not significant. Further work on these traits are needed to ascertain their role in fertilizing eggs produced by females subjected to AI. Parameters for female ostriches in the AI program were consistent with estimates from a pair-breeding flock. This is good news for the development of a viable AI program in this species, as current-flock gains are expected to selection for egg traits in female ostriches producing in the absence of males.

ACKNOWLEDGEMENTS

This research was partly funded by the Western Cape Agricultural Research Trust, the South African Ostrich Business Chamber and the THRIP initiative of the National Research Foundation. We are also indebted to Miss Naomi Serfontein for assistance with the husbandry of the flock.

REFERENCES

Bonato, M., Brand, Z., Malecki, I.A. and Cloete, S.W.P. (2014a) Book of abstracts: 10th Asia Pacific Poultry Conference, 19-23 October 2014. International Convention Centre, Jeju. P 491.

Bonato, M., Malecki, I.A., Rybnik-Trzaskowska, P.K., Cornwallis, C.K. and Cloete, S.W.P. (2014b) *Anim. Reprod. Sci.* **151**: 49.

Bonato, M., Rybnik, P.K., Malecki, I.A., Cornwallis, C.K. and Cloete, S.W.P. (2011) *Anim. Reprod. Sci.* **123**: 258.

Bonato, M., Smith, M., Malecki, I.A. and Cloete, S.W.P. (2014c) Book of abstracts: 10th Asia Pacific Poultry Conference, 19-23 October 2014. International Convention Centre, Jeju. P 492.

Bunter, K.L. and Cloete, S.W.P. (2004) Livestock Production Science **91**: 9.

Cloete, S.W.P., Engelbrecht, A., Olivier, J.J. and Bunter, K.L. (2008) *Aust. J. Exp. Agric.* **48**: 1247.

Cloete, S.W.P. and Malecki, I.A. (2011) In: The welfare of farmed ratites. Editors: P. Glatz, C. Lunam and I. Malecki. Springer-Verlag, Berlin. ISBN 978-3-642-19297-8. pp 13.

Fair, M.D., Van Wyk J.B. and Cloete, S.W.P. (2011) *S. Afr. J. Anim. Sci.* **41**: 45.

Gilmour A.R., Gogel B.J., Cullis B.R., Thompson R. (2006) ASReml User Guide. Release 2.0. (VSN International Ltd, Hemel Hempstead, HP1 1ES: UK).

Malecki, I.A., Rybnik, P.K. & Martin, G.B. (2008) *Aust. J. Exp. Agric.* **48**: 1284.

Malecki, I.A. and Rybnik-Trzaskowska, P.K. (2011) In: The welfare of farmed ratites. Editors: P. Glatz, C. Lunam and I. Malecki. Springer-Verlag, Berlin. ISBN 978-3-642-19297-8. pp. 45-64.

Rybnik, P.K., Horbanczuk, J.O., Naranowicz, H., Lukaszewicz, E. and Malecki, I.A. (2007) *Brit. Poult. Sci.* **48**: 635.

Suarez, S. and Pacey, A. (2006) *Hum. Reprod. Update* **12**: 23.

Verma, O., Kumar, R., Kumar, A. and Chand, S. (2012) *Vet. World* **5**: 301.

Proc. Assoc. Advmt. Breed. Genet. 21: 401-404

GENOMIC PREDICTIONS FOR MEAT COLOUR TRAITS IN NEW ZEALAND SHEEP

L.F. Brito[1,2], S.P. Miller[1], M.A. Lee[3], D. Lu[1], K.G. Dodds[1], N.K. Pickering[4], W.E. Bain[1], F.S. Schenkel[2], J.C. McEwan[1], S.M. Clarke[1]

[1]AgResearch, Invermay Agricultural Centre, Mosgiel, 9053, New Zealand
[2]Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, N1G 2W1, Canada
[3]Department of Mathematics and Statistics, University of Otago, Dunedin, 9058, New Zealand
[4]Focus Genetics, Napier, 4110, New Zealand

SUMMARY

The aim of this study was to evaluate the accuracy of genomic prediction for lamb meat colour traits in New Zealand sheep. A total number of 7,602 animals born between 2010 and 2013 were genotyped with the High-Density Ovine BeadChip containing 606,006 single nucleotide polymorphisms. The traits included in this study were: loin redness (A24), yellowness (B24) and lightness (L24) measured 24 hours after blooming. The significance of the fixed effects and covariates were determined using general linear model. The final fixed effects models included contemporary group, sex and birthday deviation from the contemporary group mean as a covariate. The residual from the above model was used as phenotype for the genomic evaluation model development. The software GEBV was used to calculate direct genomic values (DGV), using GBLUP methodology. To evaluate the accuracy of genomic prediction, two sets of animals were formed based on birth year: training (birth years: 2010, 2011 and 2012) and validation (birth year: 2013) sets. The accuracies for the three traits ranged from 0.29 to 0.33. Even though the accuracies were low, considering the costs and difficulty to measure and to select for meat quality traits, genomic selection might be a viable alternative.

INTRODUCTION

Meat colour traits have high economic relevance for the sheep industry as it is directly related to the appearance of the product, being an indicator of freshness and quality to consumers. In order to achieve consumer satisfaction, good management practices and environmental conditions for the animals and subsequently during meat processing play a very important role in the meat colour traits. However, it is known that meat colour has also a genetic component (e.g. Payne *et al*. 2009; Mortimer et al. 2014) and the gains achieved through genetic selection are permanent and cumulative. In order to genetically improve these traits, phenotypes must be recorded. However, they can only be recorded post-slaughter and are expensive to measure. Progeny testing not only increases the cost but also the generation interval. One possible solution to genetic improvement of these traits is genomic selection (Meuwissen *et al*. 2001), which has become a routine procedure because it reduces both progeny testing costs and generation interval. It refers to selection based on genomic breeding values, calculated from high density marker data. The benefits from genomic selection depend on the generation of accurate genomic breeding values (GEBVs). GEBV accuracies have been published for many traits in sheep (e.g. Daetwyler *et al*. 2012; Auvray *et al*. 2014). However such estimates are scarce for meat colour traits. The aim of this study was to evaluate the accuracy of genomic prediction for meat colour traits in New Zealand sheep, using the GBLUP methodology.

MATERIAL AND METHODS

Phenotypic data. Pedigree and performance records were obtained from the Sheep Improvement Limited database (SIL, www.sil.co.nz). A total of 7,602 animals born between 2010 and 2013 in the FarmIQ Progeny Test flocks (www.farmiq.co.nz) were included in this study. These animals were primarily progeny from terminal sire composites and Texel mated to a variety of maternal breeds. Animals were randomly selected to be slaughtered on given dates at commercial abattoirs. There were four to five slaughters per year and processing procedures and times were kept the same for each slaughter. The traits included in this study were: loin redness (A24), yellowness (B24) and lightness (L24) measured at 24 hours after blooming.

On the day after the slaughter, the boneless loins were vacuum packed and stored at -1°C for 8 weeks (to simulate the period that takes for chilled lamb to reach the retail market). At 8 weeks post-processing, loin pH was measured on the *Longissimus dorsi* muscle and three 2-cm thick slices of the loin were placed on small plastic trays and wrapped using semi permeable cling film and stored at 4°C (to simulate retail display) for colour measurements at 24, 48, 96 and 168 hours (seven days). Measurements were taken using a Minolta Chromometer (Konica Minolta Sensing, Inc., Osaka Japan). Three replicates were collected and the average value for each were analysed. The chromometer measures colour using the standard CIE L* a* b* colour variables (CIE L* = lightness/darkness; CIE a* = redness/brownness; CIE b* = yellowness). Only measurements at 24 hours are presented in this paper and are taken to represent the maximum redness (A value) post blooming.

The significance of the fixed effects and covariates were determined using the general linear model (GLM) procedure of SAS (SAS Inst., Inc., Cary, NC). The final fixed effects models included contemporary group, sex and to offset the differences in age of measurement, birthday deviation from the mean of the contemporary group was used as a covariate in the analysis. Contemporary group was defined by flock, birth year, sex, weaning mob (management group) and trait measurement/slaughter mob. The residual for each animal after adjusting for the above effects was used as the phenotype for the GEBV model development.

Genomic data. Marker genotypes were obtained using the Illumina High-Density Ovine BeadChip (Illumina Inc., San Diego, CA, USA), containing 606,006 single nucleotide polymorphisms (SNPs). SNPs were excluded from the analysis if minor allele frequency was less than one percent, call rate less than 90%, if they were located on the sex chromosomes, did not have known chromosome and/or position on the genome, had duplicated map positions (2 SNPs with the same position but with different names) or an extreme departure from Hardy Weinberg equilibrium ($p < 10^{-15}$). A total of 519,186 SNPs were retained for further analyses after filtering.

The software GEBV (Sargolzaei *et al.* 2009), was used to calculate direct genomic values (DGV), using the GBLUP methodology. The following model was used in genomic analysis:

$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{W}\boldsymbol{a} + \boldsymbol{e}$, where $\mathbf{y}$ is the vector of phenotypes adjusted for fixed effects, $\boldsymbol{\mu}$ is the overall mean, $\mathbf{a}$ is the vector of random animal DGVs, $\boldsymbol{e}$ is the vector of random residual effects, $\mathbf{1}$ is a vector of 1s and $\mathbf{W}$ is the design matrix linking records to animal DGVs. The DGVs were assumed normally distributed with mean zero and variance equal to $\mathbf{G}*\boldsymbol{\sigma}_{\boldsymbol{g}}^{\mathbf{2}}$, where $\mathbf{G}$ is the genomic relationship matrix based on the SNP markers and $\boldsymbol{\sigma}_{\boldsymbol{g}}^{\mathbf{2}}$ is the genetic variance. The random residual effects were assumed normally distributed with mean zero and variance equal to $\mathbf{I}*\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$, where $\mathbf{I}$ is an identity matrix and $\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$ is the residual variance.

To evaluate the accuracy of genomic prediction, two sets of animals were formed based on year of birth: training (birth years: 2010, 2011 and 2012) and validation (birth year: 2013) sets. For each trait, 10 runs were performed where each time a randomly selected group of approximately 300 animals born in 2013 (validation set) were taken as the validation set and all the animals from

the training. The youngest cohort of animals were used in validation to mimic what would happen in practice and the number of 300 was chosen in order to keep a practical number of animals in the validation set. The genomic prediction accuracy in the validation set was calculated as the Pearson correlation between DGV and adjusted phenotypes divided by the square root of heritability. The heritability was estimated from the same dataset using Restricted Maximum Likelihood (REML) procedures fitting an animal model and the same fixed effects described before, using ASReml 3.0 (Gilmour *et al.* 2009).

RESULTS AND DISCUSSION

Number of observations, trait means (± SD), trait range, coefficient of variation and the estimated heritabilities (± SE) are given in Table 1. The traits included in this study had low heritability estimates. Despite the variation in heritabilities, the DGV accuracy estimates were similar and ranged from 0.29 to 0.33 (Table 2). L24 presented the highest Pearson correlation between DGVs and adjusted phenotypes; however B24 presented a lower heritability and consequently the accuracy estimates were similar (0.32 and 0.33, respectively).

Table 1. Trait statistics and heritability estimates

Trait[1]	N	Mean ± SD	Range	CV (%)	h^2 ± SE
A24	7,602	16.79 ± 2.51	9.62 – 24.44	14.98	0.17 ± 0.03
B24	7,601	12.82 ± 2.65	5.68 – 20.31	20.68	0.11 ± 0.02
L24	7,601	40.45 ± 3.43	29.09 – 51.25	8.47	0.16 ± 0.03

[1]A24: redness/brownness; B24: yellowness; L24: lightness/darkness; N=number of animals; SD=standard deviation; CV=coefficient of variation; h^2=heritability; SE= standard error.

One reason that may contribute to the moderately low accuracies may be the low heritabilities of the traits. The accuracy estimates presented are global accuracies and robust accuracies could differ in various breed subgroups. Although our reference set appears large, almost all animals are crossbreds and/or composites from a variety of breeds. According to Saatchi *et al.* (2011) the validation is sensitive to the choice of the validation sample and to the pedigree relationships between the animals contributing to the validation and training sets, and the accuracies of DGVs are dependent on the strength of genetic relationships between the training and validation sets. It highlights the importance of maintaining an approximately constant average genetic relationship between animals in the training set and younger animals available for selection. One alternative is to define training and validation sets that are more related and also to evaluate other methodologies such as genomic evaluations using a single step procedure (Misztal *et al.* 2009) that would allow including phenotypes of non-genotyped individuals in the predictions.

In general, low accuracies of genomic breeding values limit the benefit from genomic selection. However, for traits such as meat color that are difficult to improve by traditional selection, genomic selection will be an important tool. The genomic values would help to predict breeding value of young selection candidates (without their own performance). It would result to reduced costs, shorten generation intervals, and hence accelerate the rate of genetic gain. However, future investigations are needed in order to find alternatives to increase the genomic breeding values accuracies for meat colour traits.

It is also important to highlight the need for the industry to continue investing in phenotyping and genotyping animals to create and maintain good reference and validation sets to develop accurate genomic predictions.

Table 2. Number of individuals in training and validation populations and accuracy of genomic prediction for meat colour traits

Trait	N in training	N in validation	Mean accuracy[1] (± SD)
A24	5,980	1,622	0.29 ± 0.08
B24	5,979	1,622	0.33 ± 0.11
L24	5,979	1,622	0.32 ± 0.04

[1]Mean accuracy estimated via 10 groups of around 300 animals. N=number of animals; SD=standard deviation; A24=redness/brownness; B24=yellowness; L24=lightness/darkness.

CONCLUSION

Genomic selection is likely to be a valuable tool to help in the improvement of difficult to measure phenotypes and low heritability traits such as meat colour. The findings in this study show that it is possible to generate molecular breeding values for rams at an early age for selection and breeding, thus reducing both generation interval and the costs of progeny testing. Further research will help to improve the accuracies of genomic breeding values for meat colour traits.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge: FarmIQ (Ministry for Primary Industries' Primary Growth Partnership fund), for funding the work, Landcorp and Focus Genetics for undertaking the progeny test evaluation and Silver Fern Farms for processing the animals. We thank Dr. Mehdi Sargolzaei for providing the GEBV software and we would also like to recognise the significant and on-going contribution of AgResearch farm and technical staff in management, trait recording and genotyping and AgResearch and the Brazilian Government through the Science without Borders Program that provides graduate fellowship for the first author.

REFERENCES

Auvray B., McEwan J.C., Newman S.-A. N., Lee M. and Dodds K.G. (2014) *J. Anim. Sci.* **92**: 4375.

Daetwyler H.D., Swan A.A., van der Werf J.H.J and Hayes B.J. (2012). *Genet. Sel. Evol.* **44**: 33.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. *ASReml user guide release 3.0.* VSN International Ltd, Hemel Hempstead, UK, 2009.

Meuwissen T.H., Hayes B.J. and Goddard M.E. (2001). *Genetics*. **157**: 1819

Misztal I., Legarra A.and Aguilar I. (2009) *J. Dairy Sci.*, **92**: 4648

Mortimer S.I., van derWerf J.H.J., Jacob R.H., Hopkins D.L., Pannier L., *et al.* (2014) *Meat Sci.*, **96**(2 Pt B): 1016

Payne G., Campbell A.W., Jopson N.B., McEwan J.C., Logan C.M., *et al.* (2009). *Proc. NZ Soc. Anim. Prod.* **69**: 210

Saatchi M., McClure M.C., McKay S., Rolf M., Kim J., *et al.* (2011). *Genet. Sel. Evol.* **43**: 40.

Sargolzaei M., Schenkel F.S. and VanRaden P.M. (2009). in *Dairy Cattle Breeding and Genetics Meeting. **7**(1).*

Proc. Assoc. Advmt. Breed. Genet. 21: 405-408

GENOTYPE BY BIRTH OR REARING TYPE INTERACTION IN MERINO SHEEP

A. Dakhlan[1,2], N. Moghaddar[1] and J.H.J. van der Werf[1]

[1]School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia
[2]on study leave from the University of Lampung, Indonesia

SUMMARY

This study explores if there is an interaction between the genetic potential for growth in Merino lambs and their birth (BT) or rearing (RT) type. Data consisted of 3,920 singles and 4,492 twins which were the progeny of 285 sires and 5,279 dams. We found a significant sire by BT interaction with the effect accounting for 1.59% and 2.49% of the phenotypic variation for birth weight (BWT) and weaning weight (WWT), respectively. The effect was not significant for post weaning weight (PWWT), scanned fat (SF) and eye muscle depth (EMD) with sire by BT effects accounting for less than 1% of the variation in these traits. Sire by RT interaction effects were much smaller and not significant for WWT, PWWT and SF, but accounted for 1.83% of the variation in EMD, which was significant. A bivariate analysis treating phenotypes when expressed in singles and twins as two different traits resulted in genetic correlation estimates significantly lower than one, with BT having a larger effect on genotype expression than RT.

INTRODUCTION

Birth type (BT) and rearing type (RT) constitute environments that influence the early life of sheep. Animals born as singles have higher birth weight and grow faster than animals born as twins or triplets (Yilmaz *et al.* 2007). Furthermore lambs reared as singles are heavier than those reared as twins (Safari *et al.* 2007; David *et al.* 2011). Animals born as a single are more likely to have access to better nutrition in *utero* and animals reared as a single will also have access to more milk prior to weaning compared to those reared as twins.

What has not often been looked at is whether the expression of genetic merit depends on or interacts with the BT or the RT of lambs. If such an interaction exists, there could be implications for genetic evaluation as well as for breeding programs in general. It is not feasible to design breeding programs for expression of genotypes solely as single or twins. But it may be possible to predict that the expression of breeding value is more frequent in one of these classes, and this information can be used to predict progeny differences more accurately. In genetic evaluation, it may be important to account for such genotype by BT or RT interactions if they are found to be significant.

The objectives of this study were to investigate genotype by BT or RT interactions for a number of growth related traits in Merino sheep. In a linear mixed model we investigated the presence of sire by BT or RT interaction and we estimated the genetic correlations between expressions of these traits in lambs born or raised as singles or as twins and the genetic correlation among the traits. We also investigated whether the genetic correlations between growth traits differ when expressed in singles or in twins.

MATERIALS AND METHODS

Data of Merino sheep for this study was obtained from the Information Nucleus (IN) program of the CRC for Sheep Industry Innovation in Australia. Details on this program and its design are described by van der Werf *et al.* (2010). Data consisted of birth weight (BWT), weaning weight (WWT), post weaning weight (PWWT), scanned fat (SF), and eye muscle depth (EMD). WWT was measured at approximately 100 days and PWWT, SF and EMD were measured at

approximately 250 days of age. Birth weight records were available from 8,412 lambs generated from 285 sires and 5,279 dams of Merino sheep.

Mixed model analysis was used in this study using the ASREML software (Gilmour *et al.* 2009). The fixed effects in the models were birth year (6 classes, 2007-2012 with 969-1,678 lambs per year), flock (8 classes, 521-2,483 lambs per flock), and management group within flock (GRP: up to 4 classes per flock) as one contemporary group effect, as well as a BT (2 classes) x RT (2 classes) effect. The other fixed effects included were age of dam (9 classes), sex (2 classes), and age at measurement as covariate. Live weight and age at scanning were included as fixed effects for SF and EMD.

Genetic group, animal, dam, and interaction between sire and BT or RT (SxBT/RT) were fitted as random effects in a univariate animal model. The number of genetic groups was 135 and determined by strain and flock of origin. The phenotypic variance was calculated as the sum of variance components for additive genetic effect of the animal, the dam effect, the SxBT/RT effect and the residual. A pedigree file consisting of 20,010 animals from 11 generations was used to determine additive genetic relationships among animals and account for them in the analysis. It was assumed that dams were unrelated, and in the SxBT/RT interaction terms, sires were assumed unrelated as well. We used the log likelihood ratio test (LRT) to compare the full model including SxBT/RT with a reduced model to test the significance of the SxBT/RT interaction effect.

In bivariate analyses, we considered the expression of a particular trait expressed in either singles or twins as two different traits with a genetic correlation between them (Falconer, 1952). The magnitude of the genotype by environment interaction (GEI) was evaluated based on the value of the estimated genetic correlation. Sire models were used in bivariate analyses with genetic group and sire as random effect and it was again assumed that sires were unrelated. Because of limited data in other subclasses, only traits expressed in the BT/RT combinations 11, 21, and 22 were used in the bivariate analyses to investigate GEI in an attempt to disentangle the effects of BT and of RT.

RESULTS AND DISCUSSION

Univariate analysis. Estimates of heritability (Table 1) in a model considering SxBT were lower than those without inclusion of SxBT for BWT and WWT. Heritability estimates for PWWT, SF and EMD were similar with and without inclusion of SxBT. When including SxRT in the model the heritability estimate changed only for EMD. Heritability estimates in this study are in the same ballpark as previous report (Safari and Fogarty, 2003; Safari *et al.* 2005; Mortimer *et al.* 2010).

The SxBT effect explained 1.59% and 2.49% of the phenotypic variance of BWT and WWT, respectively, which was significant and 0.76%, 0.80% and 0.06% for PWWT, SF and EMD, which was not significant. Brown *et al.* (2009) reported a similar pattern with inclusion of sire by flock-year interaction in a model, which explained 2%, 3% and 4% of variation of WWT, PWWT and yearling body weight of lambs, respectively, reducing heritability estimates by up to 50%. Maniatis and Pollott (2002) reported a similar pattern when including sire by flock-year interaction in a model, explaining only 2 to 3% of the phenotypic variation in 8 week weight and scanning weight of lambs. This result of sire x contemporary group effect explaining 2.4% of variation in body weight is similar to that reported by Pollott and Greeff (2004). The interaction term in their study explained 2% and 2% to 4% of EMD and SF variation, respectively, and heritability estimates deflated by up to 50% after accounting for GEI.

In our analysis, maternal effect contributed significantly to BWT and WWT variation (31 and 23%), but it was smaller (10%) for PWWT and these figures were very similar with and without including the SxBT effect in the model. The contribution of dam effect was also similar when

including SxRT in the model. Overall these results indicate that it may be important to include SxBT effects in the genetic evaluation of Merino sheep, particularly for BWT and WWT.

Table 1. Estimates of variance additive genetic, maternal and sire by BT(RT) effects and direct and maternal heritabilities of Merino sheep growth traits based on univariate analysis

Traits	*Variance components without SxBT or SxBT in the model**						
	σ_a^2	σ_m^2	$\sigma^2_{SxBT(RT)}$	σ_e^2	h^2	m^2	LRT
BWT	0.141	0.182		0.268	0.24 ± 0.04	0.31 ± 0.02	
WWT	1.909	2.839		7.358	0.16 ± 0.03	0.23 ± 0.02	
PWWT	7.673	2.850		16.702	0.28 ± 0.04	0.10 ± 0.02	
SF	0.087			0.283	0.23 ± 0.04		
EMD	1.590			4.074	0.28 ± 0.04		
	Variance components with SxBT in the model						
BWT	0.125	0.185	0.009	0.272	0.21 ± 0.04	0.31 ± 0.02	7.16
WWT	1.389	2.942	0.302	7.491	0.11 ± 0.03	0.24 ± 0.02	16.20
PWWT	7.207	2.673	0.205	16.856	0.27 ± 0.05	0.10 ± 0.02	1.62
SF	0.080		0.003	0.287	0.22 ± 0.04		0.67
EMD	1.582		0.005	4.078	0.28 ± 0.05		0.00
	Variance components with SxRT in the model						
WWT	1.898	2.842	0.004	7.362	0.16 ± 0.03	0.23 ± 0.02	0.00
PWWT	7.564	2.600	0.038	16.745	0.28 ± 0.05	0.10 ± 0.02	0.06
SF	0.084		0.001	0.284	0.23 ± 0.04		0.13
EMD	1.395		0.103	4.164	0.25 ± 0.05		3.90

Note: *σ^2_a = additive genetic variance, σ^2_m = maternal variance, σ^2_{SxBT} = sire by birth type interaction variance, σ^2_{SxRT} = sire by rearing type interaction variance, and σ^2_e = residual variance; and h^2 = direct genetic heritability, and m^2 = maternal heritability

Bivariate analysis. Results of this study (Table 2) show that the genetic correlation between traits expressed in singles versus twins differed significantly from one for most traits, with BT having a slightly larger effect on genotype expression than RT. This suggests that both pre- and postnatal environments significantly affect the genotype expression of weight traits in lambs. This finding agrees with Carrick and van der Werf (2005) who found that the genetic correlation between traits expressed in extreme environments (as defined by the mean performance of a cohort) was lower for earlier growth traits of sheep. In this study the expression of PWWT (at around 250 days of age) and EMD in single BT and RT might reflect the same trait and differs only in scale from twin BT and RT with genetic correlations of 0.88 ± 0.04 and 0.89 ± 0.04, respectively. Similarly, SF with the same RT but different BT (11x21) had a genetic correlation of 0.95 ± 0.02. The expression of WWT and SF in twin BT but with different RT might be the same as in twin BT and RT with genetic correlations of 0.96 ± 0.02 and 0.92 ± 0.05, respectively. Overall, these results indicate that differences in BT and RT will influence the expression of breeding values of growth traits in Merino sheep.

Table 2. Genetic correlation between traits expressed in singles or twins (born or reared) based on sire model bivariate analysis

Type of correlation*	Traits				
	BWT	WWT	PWWT	SF	EMD
11x22	0.73±0.07	0.83±0.06	0.88±0.04	0.82±0.05	0.89±0.04
11x21		0.77±0.05	0.70±0.09	0.95±0.02	0.71±0.14
21x22		0.96±0.02	0.80±0.07	0.92±0.05	0.71±0.13

Note: *11x22 = correlation between lambs born-reared as single and lambs born-reared as twins, 11x21 = correlation between lambs born-reared as single and lambs born as twins but reared as single, and 21x22 = correlation between lambs born as twins but reared as single and lambs born-reared as twins

CONCLUSION AND IMPLICATIONS

The contribution of sire by birth type interaction to the expression of birth weight and weaning weight was significant, while the contribution of sire by rearing type was only significant for eye muscle depth. In general there was genotype by environment interaction indicating that birth and rearing type influenced the expression of traits of lambs. This study suggests that BT and RT are biologically important environments that influence the genetic potential for growth of lambs. This was the case particularly for BWT and WWT which were influenced by BT and EMD that was influenced by RT. The relatively stronger interaction for BT suggests that the prenatal environment has a larger influence on the genetic expression for growth after birth compared to the postnatal environment. These results also suggest that sires could re-rank when evaluated based on single versus twin birth or rearing type. Therefore, sire by birth or rearing type interactions should be included in models used for genetic evaluation.

REFERENCES

Brown D.J., Swan A.A., Johnston D.J. and Graser, H-U. (2009) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **19**: 48.
Carrick M. J. and van der Werf J.H.J. (2005) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17**: 248.
David I., Bouvier F., François D., Poivey J.P. and Tiphine L. (2011) *Genetics Selection Evolution* **43**: 1.
Falconer D. S. (1952) *Am. Nat.* **86**: 293.
Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) *ASReml User Guide Release.*
Maniatis N. and Pollott G.E. (2002) *Anim. Sci.* **75**: 3.
Mortimer S.I., van der Werf J.H.J., Jacob R.H., Pethick D.W., Pearce K.L., Warner R.D., Geesink G.H., Edwards J.E.H., Gardner G.E., Ponnampalam E.N., Kitessa S.M., Ball A.J. and Hopkins D.L. (2010) *Anim. Prod. Sci.* **50**: 1135.
Pollott G.E. and Greeff J.C. (2004) *J. Anim. Sci.* **82**: 2840.
Safari A. and Fogarty N.M. (2003) *Technical Bulletin* **49**: 1, NSW Agriculture, Orange, Australia.
Safari E., Fogarty N.M. and Gilmour A.R. (2005) *Livestock Prod. Sci.* **92**: 271.
Safari E., Fogarty N.M., Gilmour A.R., Atkins K.D., Mortimer S.I., Swan A.A., Brien F.D., Greeff J.C. and van der Werf J.H.J. (2007) *Aust. J. Agric. Res.* **58**: 169.
van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Anim. Prod. Sci.* **50**: 998.
Yilmaz O., Denk H. and Bayram D. (2007) *Small Ruminant Res.* **68**: 336.

Proc. Assoc. Advmt. Breed. Genet. 21: 409-412

GENETIC ANALYSES ON CARCASE CHARACTERS OF AUSTRALIAN WAGYU BEEF CATTLE

Yuandan Zhang, Bruce Tier and Robert Banks

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2350, Australia

SUMMARY

Producers of Wagyu beef have unique opportunities to assess high value markets, where carcase value is determined primarily on marbling score. To date, there have been limited genetic studies of this breed, especially in Australia. Newly developed image analysis equipment could be a useful tool to analyse carcase characters. This study estimated genetic parameters for carcase traits measured by both the AUS-MEAT method and camera image analysis. Most carcase traits were moderately to highly heritable. The genetic correlation between AUS-MEAT marbling score and the image analysis marbling trait was very high, and similarly for the two measures of eye muscle area. It was concluded that image analysis of carcase traits is a feasible basis for selection in Australian Wagyu beef cattle.

INTRODUCTION

Wagyu is a collective term for Japanese beef breeds (Japanese Black, Japanese Brown, Japanese Shorthorn and Japanese Polled), accounting for 97% of Japanese cattle (Hirooka 2014). Wagyu cattle typically exhibit high marbling levels. Genetic evaluation for Wagyu cattle has been well reported but mostly for cattle in Japan and the USA. Production of Wagyu beef cattle in Australia started in the 1990s and was initiated from frozen semen and embryos, and live animals imported from Japan via the USA. The breeders' association, the Australian Wagyu Association (AWA, http://www.wagyu.org.au/), provides BREEDPLAN genetic evaluation services to Australian Wagyu breeders. While there has been increasing interest in the market potential for Australian Wagyu beef there have to date been limited genetic evaluation studies for characters of Australian Wagyu cattle. Newly developed measurements of beef quality using imaging analysis traits of carcases have been tested in Australian Wagyu cattle on a small scale (Maeda *et al.* 2014). Maeda *et al.* (2014) compared the AUS-MEAT measure and the image analysis traits of carcase characters in Australian Wagyu cattle. The preliminary results demonstrated that the image technology is a useful tool to substitute for visual assessments of carcase characters. The aims of this study were 1) to estimate genetic parameters for carcase characters of Australian Wagyu cattle measured using these image analysis traits and conventional AUS-MEAT (AUS-MEAT Ltd 2005) visual assessment; 2) to estimate genetic correlations between a subset of traits to facilitate implementation of genetic evaluation of Australian Wagyu cattle; and 3) to examine the feasibility of using image analysis traits of carcase characters as a substitute for, or to complement, the conventional AUS-MEAT grading.

MATERIALS AND METHODS

Animals and Phenotypes. Animals evaluated in this study had carcase records and were progeny of 336 sires, with progeny per sire ranging from 1 to 153. Forty-nine sires had a single offspring and 135 had fewer than 5 progeny; 6 sires had more than 100 progeny. The phenotypes and pedigree were extracted from the Wagyu BREEDPLAN database, with ancestors tracing back to the 1960's.

---

*AGBU is a joint venture of The NSW Department of Primary Industries and University of New England.

Carcase traits. Carcase traits were measured in accordance with the AUS-MEAT grading system by certified graders, and included hot carcase weight, marbling score, rib and P8 fat depths, intramuscular fat, and eye muscle area.

Image analysis traits. The image analysis traits were obtained in two steps: 1) taking digital images of carcase cross-section from between the 5th and 6th ribs using the photography device (HK-333, Hayasaka Rikoh, Sapporo Japan, Kuchida *et al.*. 2001 and Takahashi *et al.* 2006); 2) analysing the images using image analysis software (BeefAnalyserII, Hayasaka Rikoh, Sapporo Japan) to generate a set of traits based on pixel sizes, counts and colours. In brief, the muscle in the rib eye area was identified by a border line (line width of 1 pixel), this border line was semiautomatically drawn using the image analysis software and manually corrected where necessary. The image was partitioned into lean parts in white and fat (or marbling) sections in black. Fine marbling particles (fat flecks) were defined as those with an area between 0.01 and 0.5 cm^2. The image for coarse marbling particles (or fat flecks > 1 pixel) was created by further thinning the marbling image and removing hairlines of 1 pixel wide. Details of these processes have been reported previously (Kuchida *et al.* 2001). Ten image analysis traits were analysed: 1) eye muscle area (cm^2), 2) percentage of marbling area (%), 3) percentage of the coarse fat flecks (%) or marbling coarseness index as other reported (Maeda *et al.* 2014), 4) percentage of the largest fat flecks (%), 5) percentage of the largest 5 fat flecks (%), 6) percentage of the largest 10 fat flecks (%), 7) marbling fineness index or number of fat flecks per cm^2 (count/cm^2), 8) total number of fat flecks in rib eye area, 9) average luminance of the lean rib eye area and 10) ratio of minor to major axis in rib eye area.

Statistical Models. Data were analysed using an animal model fitted with fixed effects and covariates to estimate breeding values, genetic variances and heritability for traits with reasonable numbers of records and, subsequently, genetic correlations for specific pairs of traits. The fixed effects included in the analysis for carcase weight were contemporary groups (defined by herd, original owner, management group and date of slaughter), sex, and age at slaughter and age of dam were fitted as covariates. In addition to those effects the carcase character traits (measured either through AUS-MEAT or via image analysis traits) were adjusted for carcase weight instead of age at slaughter. The model used was $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$, where $\mathbf{y}$ represents the vector of observations, $\mathbf{X}$ is the incidence matrix relating fixed effects/covariate (e.g. contemporary group, sex and age at slaughter) in $\mathbf{b}$ with observations in $\mathbf{y}$, $\mathbf{Z}$ is the incidence matrix relating the random additive genetic effects in $\mathbf{a}$ with observations in $\mathbf{y}$, and $\mathbf{e}$ is the vector of random residual effects. The random effects in the model were assumed to be normally distributed with zero mean and variances as $\mathbf{Var\ (a)} = \mathbf{A}\sigma_a^2$ and $\mathbf{Var\ (e)} = \mathbf{I}\sigma_e^2$, where $\mathbf{A}$ is the numerator relationship matrix across all animals and derived from the available pedigree information, and $\mathbf{I}$ is an identity matrix. σ_a^2 and σ_e^2 are the components of variance for additive and residual random effects, respectively. Bivariate analyses were performed for pairs of traits of importance, and with sufficient data to estimate the genetic and phenotypic correlations. These included correlations between AUS-MEAT and image analysis traits for marbling, eye muscle area and fatness measures, respectively; and between carcase weight and meat quality traits (marbling, fatness and eye muscle area). The random effects in these models were assumed to be normally distributed with zero mean and variance as $\mathbf{Var(a)} = \mathbf{G_0} \otimes \mathbf{A}$ and $\mathbf{Var(e)} = \mathbf{R_0} \otimes \mathbf{I}$, where $\mathbf{G_0}$ and $\mathbf{R_0}$ are the additive genetic and residual covariance matrices among traits respectively, and where $\otimes$ is the Kronecker product.

RESULTS AND DISCUSSION

Descriptive statistics for traits are showed in Table 1. These cattle had an average carcase weight of 417kg, AUS-MEAT marbling score of 7.5, eye muscle area of 89 cm^2 and age at slaughter of 980 days. Image analyses of carcase characters showed lower eye muscle area (63 cm^2) with a greater variation than the AUS-MEAT grading measure. An average of 27.3% of the rib eye

area was intramuscular fat, with 6.8% coarse fat flecks, 4.4% (or 5.4%) the largest 5 (or 10) fat flecks and 2.68 fine fat particles per cm^2 of rib eye area. The largest fat particle accounted for 2.8% of the rib eye area and the average total number of fat particles in the rib eye area was 990 with very large variation.

Genetic parameters. Genetic parameter estimates are shown in Table 1 for carcase AUS-MEAT traits and image analysis traits. AUS-MEAT carcase traits were moderately to highly heritable. However, heritability estimates for measures of fat content traits were low; in particular, the estimate for carcase rib fat was abnormally low, with a possible explanation being irregularity in data collection (e.g. fat trimming). The estimates for carcase weight, eye muscle area and marbling are in line with the average of estimated heritability for carcase traits reported by Oyama (2011). Heritability estimates for intramuscular fat and P8 fat were similar, and the estimate for intramuscular fat was in the range summarized by Oyama (2011).

Most image analysis traits of carcase characters were moderately to highly heritable. For example, marbling percentage, marbling particle fineness and coarseness (percentages of coarse fat flecks), and number of marbling particles were very highly heritable. The brightness of eye muscle was also moderately heritable. Image analysis traits for the shape of eye muscle area (ratio of minor to major axis) and for the percentage of the largest marbling particle had low heritability. These findings are similar to those previous reported (Osawa *et al.* 2008). The current results are similar to those reported by Maeda *et al.* (2014), but with lower standard errors.

Table 1. Descriptive statistics and genetic parameters for carcase traits (including image analysis traits and AUS-MEAT measures of carcase) of Australian Wagyu cattle

Trait	N	Mean	s.d.	V_a	V_p	h^2	s.e.
Image analysis traits							
Eye muscle area (cm^2)	2095	63.0	26.6	45.5	73.5	0.62	0.11
Marbling percentage (%)	2095	27.3	7.75	13.1	35.6	0.37	0.09
Fineness Index (count/cm^2)	1856	2.68	0.62	0.13	0.26	0.50	0.11
Fat particles (%)	2095	6.82	4.53	3.13	10.3	0.31	0.09
The largest fat particle (%)	2041	2.75	2.37	0.21	4.29	0.05	0.05
The 5 largest fat particles (%)	845	4.43	3.01	4.82	6.15	0.78	0.16
The 10 largest particles (%)	845	5.43	3.43	6.41	7.56	0.85	0.15
Total number of fat particles	845	987	702	131633	164074	0.80	0.03
Minor major axis ratio	1856	0.83	1.12	0.01	0.06	0.08	0.06
luminance of muscular part	1135	80.3	8.22	22.6	57.0	0.40	0.14
AUS-MEAT traits							
Hot standard carcase weight (kg)	5269	417	61.2	646	1363	0.47	0.07
AUS-MEAT marble score 0-9	3066	7.41	1.66	1.18	2.26	0.54	0.09
Carcase eye muscle area (cm^2)	1423	88.8	18.8	52.1	88.1	0.59	0.12
Carcase rib fat depth (mm)	1303	15.0	7.00	1.48	27.6	0.06	0.05
Carcase P8 fat depth (mm)	1860	26.8	10.6	11.1	43.6	0.25	0.11
Carcase intramuscular fat (100+%)	866	134	70.7	385	1794	0.23	0.15

Genetic correlations. The genetic correlations from the bivariate analyses are shown in Table 2. The genetic correlation between values for eye muscle area measured by image analysis traits and by the traditional method is very high (0.83), and the distributions of the two measures are very similar. The genetic correlation for the carcase marbling between AUS-MEAT measure and image analysis trait was unity. Other estimates were either small or associated with high standard errors.

Table 2 Genetic and phenotypic correlations between traits

Trait1	Trait2	N*	r_g	s.e.	r_p
Carcase Weight	Image marbling%	2095	0.21	0.17	0.02
Carcase Weight	AUS-MEAT marbling score	3066	0.32	0.13	0.04
Carcase Weight	Image eye muscle area	2095	-0.15	0.14	-0.02
Carcase Weight	Carcase eye muscle area	1423	0.10	0.19	0.01
Image marbling%	AUS-MEAT marbling score	2073	1.00	0.01	0.77
Image marbling%	Image eye muscle area	2095	-0.04	0.17	0.18
Image marbling%	Image eye muscle area	1112	0.09	0.22	0.18
AUS-MEAT marbling score	Image eye muscle area	2073	-0.26	0.14	0.16
AUS-MEAT marbling score	Carcase eye muscle area	1423	0.03	0.17	0.21
Image eye muscle area	Carcase eye muscle area	1112	0.83	0.08	0.64
Carcase intramuscular fat	Carcase P8 fat	806	-0.09	0.49	-0.11

*N is the number of records common for two traits; r_g and r_p are the genetic and phenotypic correlations. s.e. is the standard error for r_g.

CONCLUSIONS

High genetic correlations show that both image analysis and AUS-MEAT measures for marbling or for eye muscle area are essentially the same traits. Image analysis of carcase characters captures useful genetic information as evident in their moderate to high heritability estimates. Thus they may be useful for selection in Australian Wagyu beef cattle. Prior to including these image analysis traits in Wagyu genetic evaluation, it is necessary to ascertain both their economic value and their genetic correlations with other traits.

REFERENCES

AUS-MEAT Limited (2005) "Handbook of Australian Meat" 7th edn. (AUS-MEAT: Brisbane)
Hirooka H. (2014) *J. Anim. Breed. and Genet.* **131**:1.
Kuchida K., Suzuki M. and Miyoshi S. (2001) *Anim. Sci. J.* **72**: 224.
Maeda S., Grose J., Kato K. and Kuchida K. (2014) *Anim. Prod. Sci.* **54**:557.
Osawa T., Kuchida K., Hidaka S. and Kato T. (2008) *J. of Anim. Sci.* **86**:40.
Oyama K. (2011) *Animal Science Journal* **82**:367-373.
Takahashi K., Kuchida K., Hori T., Nami M. and Kotaka H. (2006) *Nihon Chikusan Gakkaiho* **77**: 501.

Proc. Assoc. Advmt. Breed. Genet. 21: 413-416

CHANGE IN ACCURACY OF ESTIMATED BREEDING VALUES FOR VIAscan LEAN-MEAT YIELD WHEN ESTIMATED FROM HALF-SIB RECORDS.

H.J.B. Craig[1], N.B. Jopson[1] and M.J. Young[2]

[1] AbacusBio Limited, PO Box 5585, Dunedin 9058, New Zealand
[2] Beef & Lamb New Zealand Genetics, P.O. Box 5501, Dunedin 9058, New Zealand

SUMMARY

In-plant sheep carcass grading systems offer ram breeders the opportunity to use carcass yield data from culls in genetic evaluations. This is a special case of two-stage selection where the second-stage measurement is not made on selection candidates in the nucleus, but on half-sib relatives that have been slaughtered. Carcass yield data from ~ 11,500 progeny of 257 sires in the Beef + Lamb New Zealand Central Progeny Test were used to investigate the loss in breeding value accuracy in this scenario. All animals had pedigree recorded, weaning weight and weight of lean in the hindleg, loin and shoulder primal cuts measured by VIAscan. Datasets were prepared where VIAscan measurements were retained for a random 5%, 10%, 20%, 30%, 40% and 50% of animals, each replicated five times. Breeding values were estimated for each dataset, along with datasets comprising 0% and 100% of VIAscan records retained. Breeding values for all individuals without VIAscan records were regressed against their breeding value in the 100% dataset to determine the loss of accuracy from having records on a proportion of half-sibs rather than their own (plus their half-sibs) records. Results indicate an asymptotic increase in EBV accuracy with increasing VIAScan records. When the proportion of animals with records reaches 50%, accuracy increases to 76.0%, 67.6% and 72.8% for VIAscan leg, loin and shoulder meat yield, respectively.

INTRODUCTION

There are limitations in selection programmes for carcass traits due to the fact that many traits can only be measured post-slaughter, or in the live animal with expensive technologies such as CT scanning (Young *et al.* 1996). Measurement at slaughter is problematic because slaughter obviously precludes the animal from being a selection candidate. Meat processors in New Zealand are using carcass grading systems like VIAscan (Hopkins *et al.* 2004) which routinely give estimates of lean meat yield for carcass primal cuts. There is potentially useful carcass data collected on ram breeders' culls at slaughter that could improve meat yield breeding values. This is a special case of two stage selection (Jopson *et al.* 2004), where the primary measures are ultrasonic eye muscle dimensions, and second-stage carcass measurements are not made on highly-ranked selection candidates in the nucleus, but on half-sib relatives sent for slaughter. The accuracy of an individual's breeding value (EBV) is influenced by, amongst other things, whether the animal or its relatives were measured for a trait, the number of relatives measured and the relationship with relatives (Falconer and Mackay, 1996). The loss of accuracy in meat yield EBVs due to half-sibs of selection candidates being measured rather than candidates themselves is not known. The aim of this research was to simulate differing proportions of culls with meat yield measurements and investigate the effect on EBVs in the individuals remaining in the flock

MATERIAL AND METHODS

A progeny test dataset comprising 11,500 progeny, from 257 sires, born between 2005 and 2013, was used. Pedigree and trait data were obtained from the Beef + Lamb New Zealand Central Progeny Test (McLean *et al.* 2006). VIAscan traits analysed were weight of lean in the hindleg, loin and shoulder primal cuts (VSLEG, VSLOIN and VSSHLD, respectively; Payne *et al.* 2009); and correlated traits fitted in the Sheep Improvement Limited (SIL) meat yield module,

namely live weight at weaning and eight months of age, ultrasonic eye muscle measurement and fat depth over the eye muscle at the 12th rib (WWT, LW8, EMA and FD, respectively; Jopson *et al*. 2009).

Slaughter progeny included all the progeny of terminal sire rams and the male progeny of dual purpose rams. Slaughter animals were drafted at monthly intervals from weaning at 12 weeks of age to achieve an 18 kg carcass weight. All lambs were slaughtered at Alliance Group Plants. and evaluated using VIAscan (Hopkins *et al*. 2004) for VSLEG, VSLOIN, VSSHLD and carcass weight (CWT). The retained ewe progeny from dual purpose sires were weighed for LW8 and ultrasound scanned to measure eye muscle area (EMA) and fat depth over the eye muscle (FD) approximately 180 days after the mean lambing date of each flock.

All slaughtered progeny in the dataset had VIAscan records. New datasets were created where VIAscan records (VSLEG, VSLOIN, VSSHLD and CWT) were set to missing for a randomly selected proportion of the dataset. Datasets were produced where only a random 5%, 10%, 20%, 30%, 40% and 50% (referred to as P5 to P50, respectively) had VIAscan records present, with each dataset replicated five times. Datasets where 0 and 100% (P0 and P100, respectively) of VIAscan measurements were present were also produced, this provides datasets with EBVs estimated from correlated traits only and a dataset where EBVs are estimated from the true measurement of every animal.

EBVs were calculated using the SIL genetic engine (Young and Wakelin 2009), which are produced using a multi-trait animal model in ASReml (Gilmour *et al*. 1999). The models included terms for age of dam, birth and rearing rank, contemporary group (flock, birth year, sex, weaning mob and slaughter mob) and birthday deviation (as a covariate) for the traits WWT, LW8, FD, EMA, VSLEG, VSLOIN and VSSHLD, respectively, and a random animal effect. Breeding values were then compiled in a master dataset, and analysed in SAS (2013) with binary indicators to identify if the animal had its own VIAscan phenotypic records included or excluded for each iteration. Animals with EBVs predicted solely from relatives (i.e. their own VIAscan records were set to missing) were then regressed against the EBVs for the same animals from the P100 dataset, to produce Pearson correlation (accuracy relative to the EBV when all animals are measured for the trait).

RESULTS

The accuracy of EBVs for the three VIAscan traits estimated when no VIAscan data were present (i.e. based solely on live weight and ultrasound scanning data) ranged from 75.9 to 78.4% relative to EBVs when all animals have VIAscan records (Table 1). The average progeny per sire with VIAscan records ranged between 2.3 and 20.0 for P5 and P50, respectively. Some sires with small numbers of progeny were not represented with VIAscan records in the P5 and P10 datasets, but effectively all sires had progeny with records present for the P20 to P50 datasets.

Table 1. Mean accuracy (r) of EBVs for animals based on a proportion of half-sib relatives having VIAscan records relative to EBVs when all animals had records.

	P0	P5	P10	P20	P30	P40	P50
VSLEG	0.784	0.793	0.808	0.835	0.851	0.863	0.872
VSLOIN	0.766	0.764	0.771	0.786	0.798	0.809	0.817
VSSHLD	0.759	0.777	0.794	0.819	0.836	0.846	0.855
Sires	NA	205	239	253	255	255	257
Progeny/sire	0	2.3	4.1	7.9	12.0	16.1	20.0

The addition of VIAscan data increased the accuracy values for all three EBVs, increasing accuracy to between 81.7 and 87.2% for the P50 datasets. The majority of the improvement in accuracy was observed in the P5 to P30 datasets, with the response appearing to be heading for an asymptote in the P50 datasets. There was very good agreement in accuracy results across replicates as can be seen for VSSHLD in Figure 1.

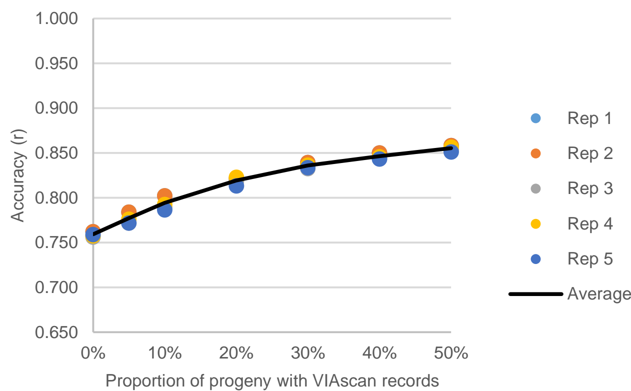


Figure 1. Accuracy of VIAscan shoulder EBVs relative to the true breeding value (TBV).

DISCUSSION

Ram breeders have a range of technologies available for prediction of meat yield to incorporate into their breeding programmes. This includes ultrasound scanning, CT scanning, and now VIAscan carcass grading as an in-plant measurement of meat yield in the hindleg, loin and shoulder primal cuts. Meat yield breeding values that use data from all three are available through SIL. However, the gains that can be made using data collected in plant have not been quantified. Some loss of EBV accuracy is expected due to the fact that the animals must be slaughtered to gather the measurement so selection candidates are not measured, compared to a situation where it is possible to measure the selection candidate itself.

The accuracy values from the P0 dataset for hindleg, loin and shoulder yield EBVs are 78.4%, 76.6% and 75.9% respectively. This level of accuracy illustrates that LW8, WWT, EMA, and FD are relatively highly correlated with the VIAscan traits (Payne *et al*. 2009) and therefore contribute significantly to the prediction of meat yield breeding values (Thompson and Meyer 1986). The results also show that a significant improvement in EBV accuracy can be made by using in-plant VIAscan measurements compared to not collecting any data at slaughter (i.e. compared to the P0 treatment). The majority of the improvement in accuracy was achieved by sampling between P20 and P30 treatments (i.e. a random 20 to 30 percent of animals slaughtered and measured). The economic benefit has not been considered, and so the cost of not having the slaughtered animals available for use in the nucleus or as rams for sale to commercial farmers has not been accounted for. This would need to be estimated to optimise the use of VIAscan measurement in a nucleus flock.

The asymptotic curve observed in Figure 1 indicates that 100% accuracy would not be achieved if it were in fact possible to collect VIAscan records on all individuals and still have selection candidates available. While this appears somewhat counter-intuitive, half-sibs only have around 25% of their alleles in common (Falconer and Mackay 1996) so 100% accuracy can never be achieved where the selection candidate is not measured for a goal trait, but some of its half-sib relatives were measured for the trait.

There are two practical factors that need to be considered in interpreting these results. Firstly, the Central Progeny Test dataset is not typical of a nucleus flock in that it is a progeny test design where a large proportion of progeny are slaughtered. In a breeder's nucleus flock, there will be many genetic connections between the rams used in the flock, all of which would be expected to improve the accuracy of breeding values, especially over time as the depth of pedigree increased (Falconer and Mackay 1996) .

Secondly, the simulated culling was based on random selection of animals. In practice farmers are likely to select a biased sample of lower-ranked animals based on live weight and ultrasonic scanning data. While some animals with high EBVs may be culled for physical faults, these are expected to be a minority. This results in truncated selection within the flock, rather than random selection. Truncated selection would also introduce unbalanced representation of sire lines in the animals sent for slaughter measurement. Thus the accuracy figure presented here may be higher than would be achieved in practice through truncated selection (Weigel 2001). However, the extent of this is unknown, but could be determined through further simulation.

CONCLUSION

Meat yield breeding values for selection candidates estimated using data collected from half-sib relatives give EBV accuracy levels intermediate between not collecting VIAscan measurements and having a technology that can measure meat yield in the live animal. Collecting measurements on a random 12 progeny per sire (P30) gives close to the maximum improvement in EBV accuracy, but an economic analysis is needed to determine the optimum level.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from Beef + Lamb New Zealand Genetics and the contribution made by Matt Parry from the University of Otago.

REFERENCES

Falconer D.S. and Mackay T.F.C. (1996) *Introduction to Quantitative Genetics.* 4th Ed. Pearson Prentice Hall.

Gilmour A.R., Cullis B.R., Welham S.J. and Thompson R. (1999) ASREML Reference Manual. NSW Agriculture.

Hopkins D.L., Safari, E., Thompson, J.M. and Smith, C.R. (2004) *Meat Sci.* **67**: 269.

Jopson N.B., Amer P.R. and McEwan J.C. (2004) *Proc. N.Z. Soc. Anim. Prod.* **64**: 212.

Jopson N.B., Newman S-A.N., McEwan J.C. (2009) *Proc. N.Z. Soc. Anim. Prod.* **69**: 158.

McLean N.J., Jopson N.B., Campbell A.W., Knowler K., Behrent M., *et al.* (2006) *Proc. N.Z. Soc. Anim. Prod.* **66**: 368.

Payne G.M., Campbell A.W., Jopson N.B., McEwan J.C., Logan C.M. and Muir P.D. (2009) *Proc. N.Z. Soc. Anim. Prod.* **69**: 210.

SAS Institute Inc. (2013) SAS/STAT 9.3 Users Guide. Cary, NC: SAS Institute Inc.

Thompson R. and Meyer K. (1986) *Livest Prod Sci* **15**: 299.

Weigel K.A. (2001) *J Dairy Sci.* **84**: E177

Young M.J. and Wakelin R. (2009) *Proc. N.Z. Soc. Anim. Prod.* **69**: 148.

Young M.J., Nsoso S.J., Logan C.M. and Beatson P.R. (1996) *Proc. N.Z. Soc. Anim. Prod.* **56**: 205.

Proc. Assoc. Advmt. Breed. Genet. 21: 417-420

EFFECT OF DIFFERENT DATA EXCLUSION POLICIES ON THE HERITABILITY OF FAT DEPTH IN MATERNAL, MERINO AND TERMINAL SHEEP

A.J. McMillan[1] and D.J. Brown[1]

[1]Animal Genetics Breeding Unit*, University of New England, New South Wales, Australia

SUMMARY

The effect of different data exclusion policies for post weaning fat depth on heritability estimates were examined using industry data for Maternal, Merino and Terminal sheep. The exclusion policies which had the most positive effect on heritability were mean fat depth for the contemporary group (CG), and standard deviation of fat depth for the contemporary group. Exclusion based on mean weight of the CG showed some effectiveness for light animals. Excluding records based on mean age of the CG had no effect on heritability estimates. In order for fat depth to be analysed most appropriately animals should be measured when they have had the opportunity to express the trait. It is appropriate to exclude CGs from the analysis where the variation in fat depth for a CG is low to improve heritability. The impact of these data exclusions on the utility of breeding values would need to be investigated in each analysis before any policies can be implemented.

INTRODUCTION

Data quality and structure impacts heritability estimates, this is particularly evident- when using industry data (Swan *et al.*2007). Various data quality factors which could have an effect on heritability estimates were examined in this study and include, mean weight of contemporary groups, age of animals in CG, mean fatness of the CG, and variability of fatness in group. In order for an animal to express its true genetic potential for fat depth the animal needs both the physical maturity and environment to deposit sufficient fat. For genetic evaluations to be able to appropriately capture that genetic merit as a breeding value the animal needs also to have contemporaries to contrast those differences.

The aim of this study was to examine the impact of data restriction policies on the heritability of post weaning fat depth using data from Maternal, Merino and Terminal sheep. The intention is also to provide feedback to the sheep industry on best practise for recording post-weaning fat and possible post measurement exclusion policies for the national genetic evaluation.

MATERIALS AND METHODS

Animals. Pedigree and phenotypes were extracted from the Sheep Genetics database (Brown *et al.* 2007). The datasets used includes 10 years (2004-2014) of records with the Terminal analyses being restricted to only Poll Dorset animals due to size constraints, while the Maternal and Merino analyses had no restriction on breed. Sheep Genetics has standard protocols to exclude fat depth records outside of 0.5 mm to 15mm, records more than 4 standard deviations from the CG mean and records with no live weight recorded at the time of measurement.

Alternate Data exclusion policies. Four data exclusion policies were developed to examine their effect on fat depth heritability estimates, and in addition an analysis with current standard data

---

* AGBU is a joint venture of NSW Department of Primary Industries and University of New England

exclusions was completed to provide a benchmark set of results. The four policies were; 1) excluding CGs where the mean weight was outside of minimum and maximum criteria, 2) excluding CGs where the mean age of the CG at recording was below age criteria and also excluding CGs where the mean age was outside of minimum and maximum criteria, 3) excluding CGs where the mean fat depth of the group was outside of minimum and maximum fat depth criteria and 4) excluding data where the standard deviation of fat depth for the group was outside of minimum and maximum standard deviation of fat depth criteria. The thresholds defined were derived by examining the distributions of these terms across the data sets.

Statistical Analysis. Genetic parameters for fat depth were estimated in univariate animal model analyses using ASReml (Gilmour et al. 2009). The fixed effects of CG and the animal's liveweight at measurement (as quadratic) were fitted. CG was defined as flock, year of birth, sex, date of measurement, management group. Random terms for the direct genetic effects and sire by flock year interactions were fitted.

RESULTS AND DISCUSSION

Heritability estimates for the Maternal, Merino and Terminal datasets are summarised in Table 1. Sheep Genetics currently use heritabilities of 0.30, 0.20 and 0.25 for post weaning fat depths in the routine evaluations for Maternal, Terminal and Merinos respectively. These benchmark heritabilities estimated are slightly lower than the current Sheep Genetics parameters for Maternal and Terminal and this reduction is likely due to the inclusion of sire by flock-year interaction term as shown by Brown et al. (2009).

Table 1. Heritability (h^2) and sire by flock-year (s^2) estimates for post-weaning fat using current standard data exclusion policies

Analysis group	Records	CGs	h^2	s^2
Maternal	143866	5609	0.24 (0.01)	0.02 (0.00)
Merino	55028	985	0.21 (0.02)	0.02 (0.00)
Terminal	249866	10014	0.21 (0.01)	0.01 (0.00)

The impact of applying the first data exclusion policy of mean CG live weight is summarised in Table 2. All analyses show a small increase in heritability from the low mean weight groups to the higher weight groups. Applying an exclusion policy solely on weight would result in improved heritability estimates, however this would be at the cost of a large proportion of data being excluded from the analysis. Heritability gains could be made if breeders waited until animals were heavier and fatter rather than measuring at lighter and leaner weight ranges.

For all breed groups the heritability estimates show no increase as age of the CG increases, using both age exclusion policies. Older groups of animals across all breeds did not show increased heritability suggesting that factors other than age are important in allowing the genetic potential for fat depth to be estimated. In all breed groups there was an increase in heritability estimates as the mean fatness of the CGs' increased (Table 3). While it would not be feasible to exclude all groups where fatness of the CG is low, breeders should be encouraged to record animals when higher levels of fatness have been achieved in the CG. Increasing the mean fatness of CG reduces the proportion of measurement error in relation to the mean .Measurement errors in fat depth scanning on average should be less than 1mm as this is a requirement of scanning accreditation (Woolcott *et al*, 2006). Heritability increases achieved through increased fatness of

CGs would allow for genetic evaluations to more accurately separate the genetic and non-genetic effects resulting in more accurate breeding values.

Table 2. Heritability (h^2) estimates for post-weaning fat depth with different contemporary group mean body weight (kg) data exclusion policies

Criteria		*Maternal*		*Merino*		*Terminal*	
Min	**Max**	**%Records**	**h^2**	**% Records**	**h^2**	**% Records**	**h^2**
30	35	N.A	N.A	9.3	0.15 (0.06)	N.A	N.A
35	40	20.4	0.23 (0.02)	22.4	0.18 (0.03)	3.3	0.06 (0.04)
40	45	25.0	0.19 (0.02)	29.6	0.14 (0.03)	11.7	0.16 (0.02)
45	50	21.2	0.20 (0.02)	19.3	0.26 (0.05)	20.3	0.17 (0.01)
50	55	15.6	0.25 (0.02)	8.4	0.26 (0.07)	21.2	0.19 (0.01)
55	60	8.0	0.32 (0.04)	7.2	0.19 (0.10)	17.0	0.17 (0.02)
60	65	N.A	N.A	1.7	0.43 (0.15)	12.2	0.22 (0.02)

N.A - Insufficient data to estimate

Table 3. Heritability (h^2) estimates for post -weaning fat depth with data exclusion based on mean fat depth of the contemporary group (mm)

Criteria		*Maternal*		*Merino*		*Terminal*	
Min	**Max**	**% Records**	**h^2**	**% Records**	**h^2**	**% Recs**	**h^2**
1	2	10.3	0.20 (0.03)	28.9	0.15 (0.03)	4.6	0.05 (0.02)
2	3	39.7	0.18 (0.01)	42.5	0.19 (0.03)	29.1	0.17 (0.01)
3	4	32.3	0.28 (0.02)	26.5	0.33 (0.03)	37.6	0.20 (0.01)
4	5	13.6	0.33 (0.03)	N.A	N.A	23.6	0.27 (0.01)

The final exclusion policy investigated was for the level of variability in fat depth within each CG (Table 4). For lower levels of fat depth variability across Maternal, Merino and Terminal analyses groups the heritability estimates were 0.10 or lower. As variability increased for all three breed groups the heritability generally increased. Thus, having greater variability in fat depth within the CGs allowed for better estimation of the genetic differences.

The factors used for data exclusion in this study are all positively correlated with each other and thus all expected to influence the level of expression in fatness. However the results confirm that the best factor to influence the heritability of fatness is to the variability of fat depth. Age and body weight provide practical proxies for breeders to assess if animals are in condition for scanning but measured fat depth is a more reliable tool.

CONCLUSIONS

These analyses show that data exclusion policies can have an effect on heritability estimates for post weaning fat depth in Maternal, Merino and Terminal sheep. Using CG based exclusion policies for mean fat depth or standard deviation of fat depth showed increases in heritability, with Maternal and Merino sheep showing the greatest benefit of these policies. Using mean weight of the group resulted in some small increases in heritability, especially when comparing lighter

groups to much heavier CGs, and could be useful for excluding very light CGs. Age at measurement had no effect on heritability estimates. In order for the most appropriate fat depth breeding values to be produced, phenotypic data should be collected for animals in CGs that have high variability for fatness. Increases in variability of CGs are indicative of groups which have exhibited the differences in their genetic potentials and provide meaningful comparisons for genetic evaluation. Excluding CGs below minimum thresholds for fat depth variability during genetic evaluation could be implemented to more accurately estimate breeding values for fat depth. However the impact such data exclusions would have on the utility of breeding values would need to be investigated and an appropriate policy developed for both historical and new data used in these analyses.

Table 4. Heritability estimates for post-weaning fat with data exclusion based on standard deviation of fat depth of each CG

Criteria		*Maternal*		*Merino*		*Terminal*	
Min	**Max**	**% Recs**	**h^2**	**% Recs**	**h^2**	**% Recs**	**h^2**
0.0	0.3	3.2	0.10 (0.07)	5.3	0.06 (0.08)	2.8	0.06 (0.05)
0.3	0.4	6.3	0.08 (0.04)	16.7	0.12 (0.04)	5.8	0.09 (0.02)
0.4	0.5	15.7	0.16 (0.02)	32.4	0.17 (0.03)	14.5	0.13 (0.02)
0.5	0.6	20.1	0.20 (0.02)	21.2	0.12 (0.04)	17.5	0.13 (0.01)
0.6	0.7	16.0	0.17 (0.02)	10.7	0.23 (0.06)	17.5	0.14 (0.01)
0.7	0.8	11.2	0.19 (0.03)	5.9	0.16 (0.05)	13.4	0.17 (0.02)
0.8	0.9	8.2	0.19 (0.04)	3.7	0.29 (0.08)	10.0	0.23 (0.02)
0.9	1.0	6.2	0.23 (0.05)	1.3	0.60 (0.20)	7.0	0.20 (0.03)

ACKNOWLEDGEMENTS

This research is funded by Meat and Livestock Australia and Australian Wool Innovation, and Sheep Genetics is made possible through the support of the Australian sheep industry. The authors acknowledge the contributions of the Sheep CRC Information Nucleus, the Australia Merino Sire Evaluation Association and industry-funded research flocks.

REFERENCES

Brown D.J., Huisman A.E., Swan A.A., Graser H-U., Woolaston R.R., *et al*. (2007) *Proc. Ass. Adv. Anim. Breed. Genet.* **17**: 187.

Brown D,J. Swan A,A. Johnston D,J. and Graser H. (2009) *Proc. Ass. Adv. Anim. Breed.* **18**: 48.

Gilmour A.R, Gogel B.J., Cullis B.R. and Thompson R. (2009) ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK

Woolcott M.,Ball A. (2005) In 'Real-time ultrasound scanning applications in livestock assessment', pp. 33-34, editor M. Wollcott and J. Allen, AGBU,University of New England, Armidale

Swan A.A, Brown D. J. (2007) *Proc. Ass. Adv. Anim. Breed. Genet.***17**: 395.

Proc. Assoc. Advmt. Breed. Genet. 21: 421-424

HOW PRECISE ARE ENTERIC METHANE EMISSION PHENOTYPES OR BREEDING VALUES ESTIMATED FROM SPOT FLUX MEASURES?

D.J. Cottle[1], J.I. Velazco[1,2], D. Mayer[3], R.S. Hegarty[1] and R.M. Herd[4]

[1]School of Environmental and Rural Science, University of New England, Armidale, NSW 2351
[2]National Institute of Agricultural Research, Ruta 8 km 281, 33000 Treinta y Tres, Uruguay
[3]Department of Agriculture and Fisheries, Ecosciences Precinct, Dutton Park Qld 4102
[4]NSW Department of Primary Industries, Beef Industry Centre, Armidale, NSW 2351

SUMMARY

This study was conducted to determine the precision or confidence interval of phenotypes and estimated breeding values for enteric methane (CH_4) emissions of individual animals estimated from spot measurements of methane flux. Components of variance of daily methane production (DMP; gCH_4/d) and methane yield (MY; gCH_4/kg DMI) from trial data were used to estimate the precision for assumed heritability values. The precision was relatively insensitive to number of measures per animal per day and to the number of days of measurement. The values of the residual components of variance (between measures, within-animals, within-days) are high compared to between animal and between day variance but the confidence intervals for EBVs for DMP and MY estimated from spot flux measures are about 20% each side of the mean, which should be adequate for industry implementation in breeding schemes.

INTRODUCTION

Cattle breeders can select for lower methane (CH_4) production directly via the use of methane measuring equipment, such as respiration chambers (Herd *et al.* 2014), Greenfeed emission monitoring units (GEM; Velazco *et al.* 2014) or indirectly via pasture feed intake (Cottle 2011, 2013). Indirect selection is only superior to direct selection if the indirect measurement is easier to make, has a high genetic correlation with the direct trait and has moderate to high heritability. At present measuring pasture intake of large numbers of cattle is no easier than measuring CH_4 production. On-farm measurement of DMP is likely to occur without knowledge of the dry matter intake, although herd intake may be determined (Jones *et al.* 2011). The simplicity of obtaining short-term (spot) measurements of enteric CH_4 production rate is causing these methods to be evaluated for their use in estimating genetic parameters for CH_4 production (Pickering *et al.* 2013). Typically, the arithmetic average of spot measures is used to estimate daily CH_4 production (DMP; g CH_4/d) yet the precision of this approach has not been reported (Cottle *et al.* 2015). Emission rates are known to change over momentary, diurnal and longer seasonal patterns (Crompton *et al.*, 2011; Ulyatt *et al.*, 2002; Munger and Kreuzer, 2008), requiring representative sampling.

This study aimed to determine the precision or 95% confidence interval of individual phenotypes and EBVs for CH_4 emissions estimated from ~3-5 minute, spot measurements of enteric methane flux.

METHODS

Two data sets (grazing and feedlot; Cottle *et al.* 2015) were used to calculate the minimum number of spot flux measures needed to phenotype the true average CH_4 emissions of an animal as required to develop DMP estimated breeding values (EBVs). DMP was estimated from multiple 3-5 min spot measures of methane flux made by the GEM system using 24 cattle. The analysis was based on an acceptable margin of error (MoE) for sampling, a level of confidence to be associated with the final estimates, and an estimated coefficient of variation for each particular sample. MoE is the maximum permitted deviation of the estimate from the true mean. These calculations assume

the confidence level for sampling would be 90% (i.e. the measured value of DMP should be within 10% of the true value). DMP estimates from cattle were estimated by GEM while cattle grazed pastures (173 gCH_4/d) then again when they were feedlot finished (DMP = 142 gCH_4/d). The MoE for each individual methane measurement was chosen as ± 5-10 gCH_4/d. Measurement errors expressed as a percentage of the means, when MoE = 10 gCH_4/d were therefore, 100*(10/142) = 7%, and 100*(10/173) = 6% for feedlot and pasture respectively. For desired margins of error and levels of confidence, sample sizes were calculated as follows:

Sample size (N) = $(z^2 * CV^2)/(MoE/\mu)^2$

where:

z is the value associated with the chosen confidence interval;

CV was 40% (feedlot) or 30% (pasture); and

MoE/μ is the ratio between the margin of error and the mean.

To determine the optimum number of days and measures per day to achieve desired precisions of phenotype estimates and EBVs for DMP and MY, the 95% two-tailed confidence interval was estimated from the feedlot variance estimates reported by Cottle *et al.* (2015). The standard error was calculated using the formulae in Cox and Solomon (2003) as shown below:

Standard error (mean) = $\text{sqrt}\ [\sigma^2 / (n_a.n_d.n_r) + \tau_a / n_a + \tau_d / n_d]$,

where:

σ^2 is the residual variance;

n_a, n_d and n_r are respectively the numbers of animals, days and samples per day, and

τ_a and τ_d are the variance components for animals and days respectively.

The confidence intervals for EBVs were estimated as $\pm 1.96 * \sqrt{((\sqrt{(1-h^2)} * \sqrt{V_A}) + V_E)}$,

where:

h^2 is heritability,

V_A is additive genetic variance, and

V_E = environmental variance $((1-h^2)*V_P)$.

RESULTS AND DISCUSSION

To be 90% confident of the DMP phenotype estimate being within 7.5% of the true mean 68 spot measures were needed from an animal in a feedlot situation and 60 spot measures under grazing conditions (Table 1). CV is lower at pasture for a given absolute MoE as DMP is higher.

Table 1. Number of short-term GEM measures required to estimate the DMP phenotype of an individual animal (g CH_4/day) with a specified margin of error and with a defined confidence using feedlot and grazing data sets

MoE (gCH_4/d)	Confidence interval (%)			
	70	80	90	95
Feedlot data set				
5	61	93	153	217
7.5	27	41	68	97
10	15	23	38	54
Grazing data set				
5	54	81	134	190
7.5	24	36	60	85
10	13	20	34	48

The relationships between 95% confidence intervals and days of measurement and measurements per day are shown for DMP and MY EBVs in Figure 1.

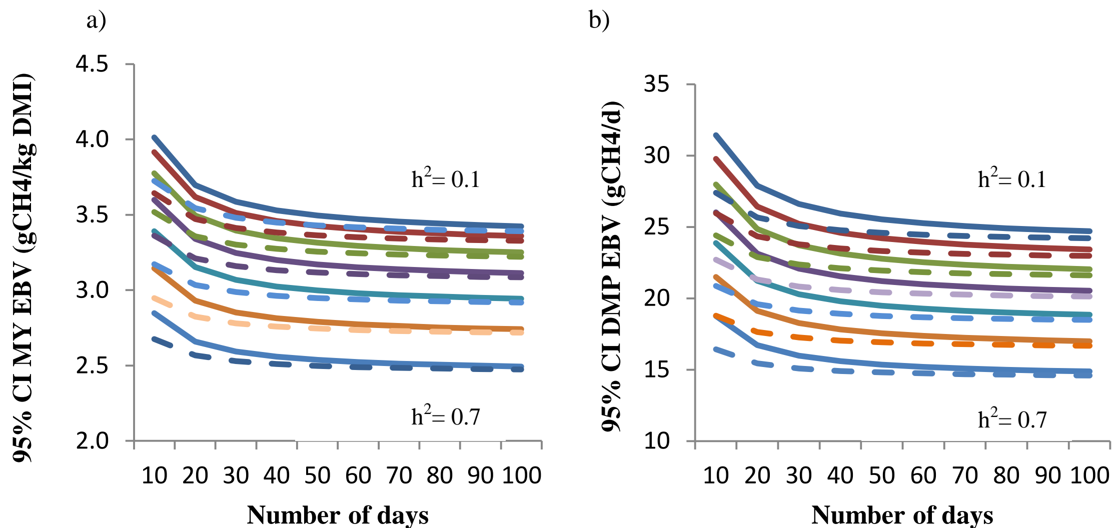


Figure 1. Estimated width of 95% confidence intervals of a) DMP EBVs and b) MY EBVs (either side of mean) vs. numbers of days with 2 measurements / animal / day (solid line) or 10 measurements / animal / day (dashed line). Heritabilites from top to bottom: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7.

The number of spot measures per animal per day (n_r) is dependent on the frequency of supplement delivery by the GEM and voluntary visitation by cattle which is largely outside the control of the researcher. Visitation had a minor effect on the number of days required to achieve a target precision for phenotypes. It would seem prudent to assume $n_r = 2$, knowing that a higher number will slightly improve precision. Measuring animals less than 50 times will probably not achieve desired phenotype MoE and confidence intervals.

The power analyses suggested that spot measurements would result in a precision in the DMP estimate of <10% deviation from true DMP value if they are made over a 70d period as routinely used with RFI tests in a feedlot. Spot measurements of enteric emissions can be used to define DMP but the number of animals and samples are larger compared to measurements made in respiration chambers with a lower CV (Hegarty 2013).

Regarding establishing a precise estimate of the long-term emission phenotype, in a feedlot an animal needed 54 spot emission measurements to be 95% confident that the estimated mean is within 10% of the true DMP phenotype (Table 1). If MoE is 7.5 g/d, the required minimum number of measures (n=60) to describe a grazing animal's phenotype within 10% of the true mean DMP, can be achieved by sampling an animal twice a day over 30 days, or 5 times a day over 12 days. The more intense sampling schedules could confound the estimates under grazing conditions because a higher amount of supplement per day is required to attract the animals into the GEM unit. Within those ranges, all combinations of sampling regimes should deliver estimates within 10% of the true phenotype. Less intense sampling regimes may increase the number of animals utilising a GEM unit.

There is a minimum data requirement for all EBV traits so the optimization of the CH_4 test duration will seek to provide the data at the lowest cost. A 35 day test was suggested by Archer *et*

al. (1997) to be sufficient to phenotype an animal's feed intake (critical for the calculation of MY). In that case, 2 flux measures per day would enable the phenotype of DMP for a specific age and animal class to be used to calculate MY. If DMP is to be related to growth rate, a minimum 70 day test length with cattle weighed every two weeks is suggested (Exton 2001) so a 70 day test for growth rate can easily be run concurrently with the CH_4 determinations.

From Figure 1, the 95% confidence interval for DMP EBV estimates was ±25 gCH_4/day and for MY EBV estimates was ±3.5 gCH_4/kg DMI, assuming a heritability of 0.26 for DMP and 0.23 for MY (Lassen and Lovendahl 2013). Increasing the number of measurements / animal / day or number of days of measurement, (i.e. total number of measurements), had little impact on the precision of EBVs. The confidence intervals are about 20% of the mean values for DMP (~150g CH_4/day) and MY (~13g CH_4/kg DMI) each side of the mean, which is a relatively wider confidence interval than most traits, but should be of adequate precision for use in industry via breeding schemes such as Breedplan. The design of future enteric CH_4 experiments will usually depend on the available budget and logistic limitations. Our formula and results can be used as a guide for any future experimental designs.

ACKNOWLEDGEMENTS

Bruce McCorkell, NSW DPI, provided the formula for calculating the sample sizes needed for phenotypes. Julius van der Werf, UNE, provided the formula for calculating the 95% CI_{EBV}. This work was supported in part by the Australian Government's "Filling the Research Gap" program.

REFERENCES

Archer J.A., Arthur P.F., Herd R.M, Parnell P.F. and Pitchford W.S. (1997) *J. Anim. Sci.* **75**: 2024.

Cottle D.J. (2011) *Proc. Aust. Assoc. Anim. Breed.* **19**: 423.

Cottle D.J. (2013) *Animal Production Science* **53**: 1209.

Cottle D.J., Velazco J.I., Hegarty R.S. and Mayer D. (2015) *Animal.* doi: 10.1017/S1751731115001676

Cox D.R. and Solomon P.J. (2003) 'Components of Variance'. Chapman & Hall / CRC, Boca Raton, USA.

Crompton L.A., Mills J.A.N., Reynolds C.K. and France J. (2011) In 'Modelling nutrient digestion and utilisation in farm animals'. pp. 176–180. editors D Sauvant, J Milgen, P Faverdin, N Friggens, Wageningen Academic Publishers: Wageningen, The Netherlands.

Exton S. (2001) http://www.dpi.nsw.gov.au/agriculture/livestock/beef/feed/publications/feed-efficiency [Verified 17 April 2015]

Hegarty R.S. (2013) *Animal* **7**, 401.

Herd R. M., Arthur P. F., Donoghue K. A., Bird S. H., Bird-Gardiner T. and. Hegarty R. S. (2014) *J. Anim. Sci.* **92**: 5267.

Jones F.M., Phillips F.A., Naylor T. and Mercer N.B. (2011) *Anim. Feed Sci. Tech.* **166**: 302.

Lassen J. and Lovendahl P. (2013) Heritability for enteric methane emission from Danish Holstein cows using a non-invasive FTIR method. Advances in Animal Biosciences **4**: 280.

Münger A. and Kreuzer M. (2008) *Aust. J. Exp. Agric.* **48**: 77.

Pickering N.K., de Haas Y., Basarab J., Cammack K., Hayes B., Hegarty R.S., Lassen J., McEwan J.C., Miller S., Pinares-Patiño C.S., Shackell G., Vercoe P. and Oddy V.H. (2013) http://www.asggn.org/publications,listing,95,mpwg-white-paper.html [Verified 15 February 2015]

Ulyatt M.J., Lassey K.R., Shelton I.D. and Walker C.F. (2002) *N.Z. J. Agric. Res.* **45**: 217.

Velazco J.I., Cottle D.J. and Hegarty R.S. (2014) *Anim. Prod. Sci.* **54**: 1243.

Proc. Assoc. Advmt. Breed. Genet. 21: 425-428

IMPLICATIONS OF MANIPULATING THE EWE LIVE WEIGHT PENALTY IN MATERNAL SHEEP INDEXES IN NEW ZEALAND

J.I. Kerslake[1], P.R. Amer[1], M.J. Young[2] and T.J. Byrne[1]

[1] AbacusBio Ltd, 442 Moray Place, PO Box 5585, Dunedin, New Zealand, 9058
[2] Beef and Lamb New Zealand Genetics, PO Box 5501, Dunedin, New Zealand, 9058

SUMMARY

Understanding the consequences of selection represents an important part of the development of a genetic improvement program. The weighting applied to the adult ewe live weight (EWT) estimated breeding value (EBV) in New Zealand dual purpose flocks is currently highly topical with breeders having contrasting views as to the appropriateness of the current strong downward selection emphasis. This research assessed the implications of selection using a restricted ewe live weight index, and quantified the loss of efficiency of selection on indexes with varying economic weights, and a zero economic weight, for the EWT EBV, in flocks recording and not recording EWT.

Results showed that recording ewe weight enables EWT EBV change to be restricted while achieving increased rates of gain in early growth traits. The current dual purpose production (DPP) index (Byrne *et al.* 2012) was also found to be robust to a 17 to 33% reduction in the EWT economic weight, resulting in a 2 to 4% loss in efficiency of selection on the current DPP index for all flocks that are either recording or not recording EWT. While selection indexes were robust to changes in EWT economic weights, if the EWT economic weight was set to zero, equivalent to a decision to exclude EWT from the breeding goal, the loss in efficiency of selection on the current DPP index was 16%.

One option for industry could be to implement a 33% reduction in EWT economic weight which would result in no reduction in genetic potential for ewe mature size while selection candidates with superior growth rate would rank more consistently with ram breeder and buyer expectations. While this compromise would typically result in a 2 to 4% loss in efficiency of the current DPP, such an outcome is preferable to exclusion of EWT from the breeding goal as currently practiced by some breeders who object to a strong negative penalty on EWT, because that strategy leads to a 16% reduction in the economic value of genetic progress.

INTRODUCTION

Understanding the consequences of selection represents an important part of the development of a genetic improvement program. It enables breeders and farmers to understand how animal performance will change over time as a result of selection. A number of sheep breeders have provided feedback to Beef + Lamb New Zealand Genetics (B+LNZ Genetics) that some of the consequences of selection on the current DPP index are not desirable. Of specific concern is the loss of gain in early growth, as a result of using a negative economic weight on the EWT EBV. Depending on the level of recording and accuracy of prediction, the current DPP index may increase or decrease ewe weight (Table 1). While it is recommended that a negative economic weight on the EWT EBV be included in the DPP index, some breeders are requesting that this be dropped in the genetic evaluation of their flocks. Moving forward breeders and farmers would like to have the ability to be able to control EWT with continual improvement in early growth.

The aim of this research was to assess the implications of selection using a restricted EWT index and to quantify the los, of efficiency, relative to the current DPP index, of selection on overall indexes with varying economic weights for the EWT EBV, in flocks recording and not recording EWT.

MATERIALS AND METHODS

Data collation. The Sheep Improvement Limited Database (Newman *et al.* 2000) was used to identify 79 flocks of different breeds (Romney, Perendale, Coopworth and Composite) that were recording or not recording EWT from 2006-2013. This data was collated into high and low EWT BV accuracy datasets from flocks recording or not recording EWT. A quantification of the expected loss of efficiency of selection on indexes with varying economic weights for EWT was undertaken (Table 1).

Data analysis. The first step of the approach involved estimating the regression coefficients of each trait of interest on the index (DPP) in question, within high and low accuracy datasets from flocks recording or not recording EWT. These regression coefficients (b), interpreted as how many units of progress in a trait can be expected per unit change in the index, can be derived from genetic variances of traits and indexes, as follows:

$$b_{T,I} = r_{T,I} \times \sqrt{\left(V_T / V_I \right)} \quad \{\text{Equation 1}\},$$

where r is the correlation and V is the variance for trait T and index I, respectively. These values are very simply calculated for any set of selection candidates which have EBVs for the traits of interest, and for any specified index.

The next step is to set as a benchmark the rate of genetic progress being achieved using the current index. This can be evaluated by looking at the averages of index values for animals born by birth year over recent years, a standard and routine practice in most genetic evaluation systems. If we assume that the vast majority of genetic progress comes from selection of a single type of selection candidate (e.g. progeny tested sires), then response to selection (R) on the current index (IC) is:

$$R_{IC} = i \times r_{IC,TM} \times \sqrt{(V_{TM})} / L \quad \{\text{Equation 2}\},$$

where i is the selection intensity, r is the accuracy of selection of candidates on the current index (IC), V is the variance for true overall merit (TM), and L is the generation interval. If we assume that selection intensity and generation interval will be the same irrespective of what index is used to achieve genetic progress (this is reasonable for similar indexes with just moderately modified weightings on the same or similar traits as in the current index), then the relative rates of response in two indexes will be:

$$R_{IN} / R_{IC} = r_{IN,TM} \times \sqrt{V_{TM}} / r_{IC,TM} \times \sqrt{V_{TM}} \quad \{\text{Equation 3}\},$$

where IN is the new index, and the other parameters are described in equation 2. We can predict response in any trait of interest (R_T) resulting from selection on the new index so that 100 units of index progress is achieved based on combining Equations 1 and 2 above to be:

$$R_T = 100 \times r_{T,IN} \times \sqrt{V_T / V_{IC}}$$

It is important to note that in the above calculation, the correlation described is for the trait of interest with the new index (IN) being considered, while the variance of the current index (IC) is used to standardise the results in the denominator of the equation.

The equations can be easily adapted to predict response in one index that arises from selection on another index. This is achieved by treating the index of interest in the same way as we treated the trait of interest as described above.

RESULTS AND DISCUSSION

Results show that when selection candidates have high accuracy for the EWT eBV, selection on the current DPP index is expected to result in a modest reduction in EWT (Table 1). Table 1 also shows that the recording of EWT enables the EWT EBV change to be restricted while achieving increased rates of gain in early growth traits (e.g. for high accuracy recorded flocks, CW response increases from 0.058 to 0.071, when EWT is restricted to zero). A similar result was observed for flocks that are not recording EWT, although the realised increases in rates of gain in early growth rate are much smaller (e.g. for high accuracy non recorded flocks, CW response increases from 0.067 to 0.068). This is because an almost equivalent weight (-146) to the current weight (-149) is required to restrict EWT EBV gain in flocks not recording EWT (Table 1).

Table 1. Expected responses (regression coefficients) for each trait and index if selecting for 100 cents of progress in the current index.

Dataset		**High accuracy**				**Low accuracy**			
Ewe live weight (EWT)		**Recorded**		**Not recorded**		**Recorded**		**Not recorded**	
Selection candidates		**n=247,840**		**n=69,725**		**n=28,333**		**n=63,085**	
Accuracy of EWT eBV		**67.6%**		**63.5%**		**49.7%**		**50.3%**	
Economic weight for EWT eBV (cents)[1]		**Cur -149**	**New -109**	**Cur -149**	**New -146**	**Cur -149**	**New- 177**	**Cur -149**	**New -210**
eBV[2]	**unit**								
NLB	lamb	0.011	0.011	0.012	0.012	0.008	0.008	0.010	0.010
SUR	lamb	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
SURM	lamb	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WWT	kg	0.094	0.113	0.126	0.128	0.143	0.125	0.165	0.129
WWTM	kg	0.050	0.053	0.044	0.044	0.079	0.075	0.060	0.051
CW	kg	0.058	0.071	0.067	0.068	0.084	0.072	0.087	0.063
EWT	kg	-0.067	0.000	-0.006	0.000	0.047	0.000	0.095	0.000
FW12	kg	0.012	0.014	0.013	0.013	0.019	0.018	0.016	0.012
LFW	kg	0.002	0.002	0.002	0.002	0.003	0.002	0.002	0.002
EFW	kg	0.011	0.012	0.011	0.012	0.017	0.015	0.014	0.011
Index[3]	**unit**								
DPP (IC)	¢	100.0	98.63	100.0	99.99	100.0	99.33	100.0	96.99
DPPR (IN)	¢	97.31	98.64	99.98	99.99	98.68	99.33	94.19	97.03

[1] Cur: Ewe live weight economic weight in the current index; New: Ewe live weight economic weight required to restrict change in ewe live weight. [2] eBV, Estimated breeding values; NLB, number of lambs born; Sur, survival; SurM, survival maternal; WWT, weaning weight; WWTM, weaning weight maternal; CW, carcase weight; EWT, ewe live weight; FW12, fleece weight at 12 months; LFW, lamb fleece weight; EFW, ewe fleece weight. [3]Index: DPP (using current ewe live weight economic weight); DPPR, dual purpose restricted to zero change in EWT eBV (using the restricted ewe live weight economic weight)

When the accuracy of prediction of genetic merit for EWT EBV is low, significantly higher economic weights are required to restrict genetic change in EWT, for recording (-177) and (-210) non recording flocks (Table 1). With this level of weighting, associated reductions in response to

selection for early growth traits are apparent. The magnitude of this reduction (from responses to selection in current index) is greater for non-recording flocks compared to recording flocks. This shows an inability to identify genetic variation for adult ewe weight independent of early growth, when the trait is not recorded or predicted with low accuracy.

As theory defines, results showed that selection indexes are robust to modest changes in the economic weight for EWT (Figure 1). For example if the current economic weight is dropped or increased by 50%, 95-96% of the current DPP will still be realised for flocks that are either recording or not recording ewe live weight. This efficiency increases to above ~99% with a 17% increase or decrease in the economic weight for EWT.

If the EWT penalty is set to zero, equivalent to the practice of dropping EWT from the breeding goal, the loss in efficiency of the current DPP was 16%. To encourage industry to keep EWT in the breeding goal, the EWT penalty could be dropped by 33%, resulting in no reduction in genetic potential for ewe mature size when accuracy is high in flocks measuring the trait while selection candidates with superior growth rate would rank more consistently with ram breeder and buyer expectations. While this compromise could result in a 2 to 4% loss in efficiency of selection on the current DPP index for all flocks that are either recording or not recording EWT, this is a far better outcome for industry than a 16% loss from dropping EWT from the breeding goal.

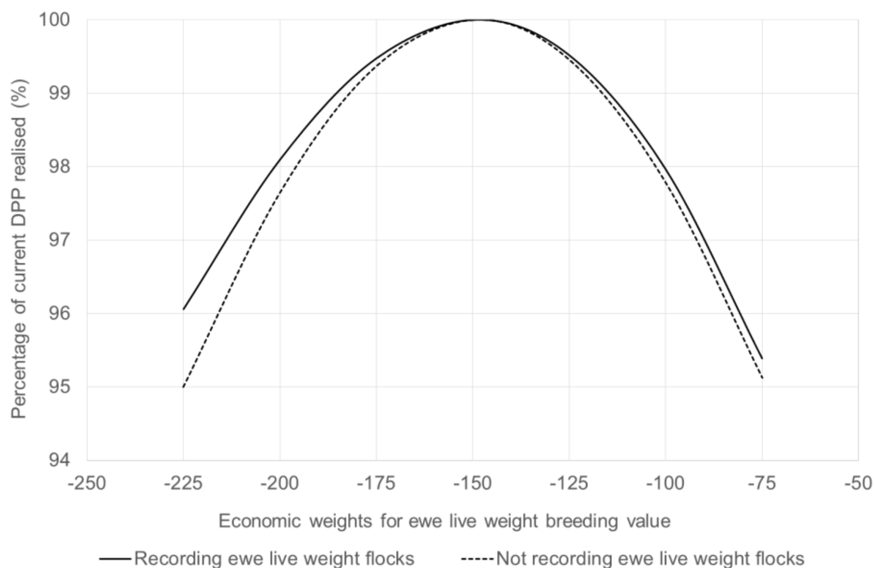


Figure 1. The effect of different economic weights for adult ewe live weight estimated breeding values on the percentage of current Dual purpose production index (DPP) realised for selection candidates from flocks that are recording or not recording ewe live weight.

REFERENCES

Byrne T.J., Ludemann C.I., Amer P.R., Young, M.J., 2012. *Livestock Science* **144**: 20.

Newman S.A., Dodds K.G., Clarke J.N., Garrick D.J., McEwan J.C., 2000. *Proc. New Zealand Soc. Ani. Prod.* **60**: 195.

Proc. Assoc. Advmt. Breed. Genet. 21: 429-432

COMPUTING FOR MULTI-TRAIT SINGLE-STEP GENOMIC EVALUATION OF AUSTRALIAN SHEEP

Karin Meyer, Andrew Swan and Bruce Tier

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

SUMMARY

The impact of parameterising to genetic principal components and dimension reduction on computational requirements is examined for a subset of traits considered in single step evaluation of sheep in Australia. Together with judicious treatment of dense blocks due to genomic relationships in the mixed model equations, such models can reduce computational requirements many-fold.

INTRODUCTION

Genetic evaluation utilizing genomic information is in the process of being adopted in many livestock improvement schemes. In particular, the so-called 'single-step' procedure allows for joint evaluation of all animals – genotyped or not – utilising all pedigree information available at the same time (Misztal *et al.* 2009). It can be thought of as an extension of previous, best linear unbiased prediction schemes, replacing the pedigree based numerator relationship matrix between animals, **A**, by it's counterpart, **H**, which combines the genomic relationship matrix among genotyped animals, **G**, with relationships derived from the pedigree. Computing the inverse of **H** requires large matrix products and direct inversion of **G** and the corresponding submatrix of **A**, and challenges thus posed have attracted considerable attention (e.g. Aguilar *et al.* 2011).

Computational requirements to estimate breeding values are heavily dependent on the number of non-zero (NNZ) elements in the coefficient matrix, **C**, of the mixed model equations (MME) to be solved. In a multivariate analysis comprising q traits, each non-zero element of the inverse of the relationship matrix can contribute up to q^2 elements to **C**. Equivalent and reduced rank models have been proposed which can reduce this number (Meyer and Kirkpatrick 2005; Meyer 2009), but have seen little practical use. Let animals be grouped according to their genomic information status, with $\mathbf{H}_{22}$ the submatrix of **H** for genotyped individuals. Typically, $\mathbf{H}_{22}$ and the corresponding block of $\mathbf{H}^{-1}$ are dense, i.e. contain few zero elements. Hence, the NNZ elements in **C** arising from elements of $\mathbf{H}^{-1}$ becomes more important than previously, where the inverse relationship matrix $\mathbf{A}^{-1}$ was sparse throughout. Furthermore, existence of dense blocks in the MME together with substantial amounts of random access memory (RAM) available in modern hardware readily allow matrix manipulation routines from highly optimized software libraries to be exploited. We examine the utility of equivalent or reduced rank models together with the use of multi-threaded library routines for dense matrix calculations for an application of single-step genetic evaluation to Australian sheep data.

EQUIVALENT MODELS AND BEYOND

Consider a linear mixed model for q traits, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}$ with $\mathbf{y}$, $\boldsymbol{\beta}$, $\mathbf{u}$ and $\mathbf{e}$ the vectors of observations, fixed and random effects, and residuals, and **X** and **Z** the pertaining incidence matrices. Let **u** represent animals' additive genetic effects, ordered by animals within traits so that $\text{Var}(\mathbf{u}) = \boldsymbol{\Sigma} \otimes \mathbf{H}$, with $\boldsymbol{\Sigma}$ the genetic covariance matrix among traits. For $\text{Var}(\mathbf{e}) = \mathbf{R}$, the diagonal block in **C** for **u** is then $\mathbf{C}_{uu} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \boldsymbol{\Sigma}^{-1} \otimes \mathbf{H}^{-1}$. The first part, $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$, is block-diagonal for animals, with blocks of size $q \times q$. If $\boldsymbol{\Sigma}^{-1}$ has no zero elements, $\boldsymbol{\Sigma}^{-1} \otimes \mathbf{H}^{-1}$, however, contributes q^2 non-zero elements to $\mathbf{C}_{uu}$ for each non-zero off-diagonal element of $\mathbf{H}^{-1}$.

An equivalent model is obtained by expanding **Zu** to $\mathbf{Z}(\mathbf{Q} \otimes \mathbf{I})(\mathbf{Q}^{-1} \otimes \mathbf{I})\mathbf{u} = \mathbf{Z}^{\star}\mathbf{u}^{\star}$, with **I** an

*AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

identity matrix. This gives $\text{Var}(\mathbf{u}^\star) = \mathbf{Q}^{-1}\boldsymbol{\Sigma}\mathbf{Q}^{-T} \otimes \mathbf{H} = \boldsymbol{\Sigma}^\star \otimes \mathbf{H}$ and $\mathbf{C}^*_{uu} = \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{Z}^\star + (\boldsymbol{\Sigma}^\star)^{-1} \otimes \mathbf{H}^{-1}$. Choosing $\mathbf{Q}$ so that $\boldsymbol{\Sigma}^\star$ is diagonal reduces the NNZ elements contributed by each non-zero element of $\mathbf{H}^{-1}$ to q. The trade-off for this is that $\mathbf{Z}^\star$ has up to q non-zero elements per observation compared to, typically, a single element of unity in $\mathbf{Z}$. This gives rise to some extra non-zero elements in other parts of $\mathbf{C}^\star$, especially in the off-diagonal block for fixed × random effects, $\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}^\star$. Suitable matrices $\mathbf{Q}$ can be obtained from the eigen-decomposition $\boldsymbol{\Sigma} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}'$, either the matrix of eigenvectors, $\mathbf{Q} = \mathbf{E}$, or the matrix of 'factor loadings', $\mathbf{Q} = \mathbf{E}\boldsymbol{\Lambda}^{-1/2}$. The latter can be rotated to lower triangular form, $\mathbf{Q} = \mathbf{E}\boldsymbol{\Lambda}^{-1/2}\mathbf{T}$ (with $\mathbf{T}\mathbf{T}' = \mathbf{I}$) which reduces the NNZ elements in $\mathbf{Q}$ to $q(q+1)/2$ and thus the number of multiplications to set up the MME and the NNZ in $\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}^\star$.

Furthermore, this parametrization can directly yield substantial, additional computational savings by invoking a 'reduced rank' model, if $\boldsymbol{\Sigma}$ has $q - r$ negligible eigenvalues, which generally holds for larger values of q. This involves estimating only the first r principal components (i.e. elements of $\mathbf{u}^\star$) for each animal which, at convergence, are combined to give the q corresponding elements of $\mathbf{u}$. This is achieved by simply considering only the first r columns of $\mathbf{Q}$, which reduces the number of equations in the model as well as the NNZ elements.

MATERIAL AND METHODS

Data consisted of 5.24 million records for 5 traits recorded on 1.77 million animals in the LAMBPLAN terminal sire breeds evaluation (Brown *et al.* 2007), representing the most commonly recorded traits in these breeds, namely birth, weaning and post-weaning weights, and post-weaning eye muscle and fat depth. Including parents without records there were 1,995,755 animals of which 10,698 (N) were genotyped for 48,599 single nucleotide polymorphisms. To build $\mathbf{H}^{-1}$, genomic relationships were computed following Yang *et al.* (2010). This yielded 63,793,942 NNZ elements in $\mathbf{H}^{-1}$ (halfstored), compared to 6,584,393 elements in the corresponding pedigree based matrix $\mathbf{A}^{-1}$.

As in the routine LAMBPLAN evaluation, records were pre-corrected for the effects of birth-rearing type, age at measurement and age of dam, and body weight as a covariate for eye muscle and fat depth. The model of analysis then comprised contemporary groups as fixed effects, animals' additive genetic effects, dams' permanent environmental effects for the body weights (653,067 levels), and genetic groups (93 levels) as random effects. The latter were fitted 'explicitly' – assigning proportions of membership for each animal – as augmenting the pedigree by phantom parents in single-step applications can be problematic (Misztal *et al.* 2013).

Analyses fitted standard multivariate (MV) and the principal components (PC) models described above. Dense diagonal blocks in $\mathbf{C}$ (or $\mathbf{C}^\star$) for genotyped animals were stored in two-dimensional arrays, a single matrix of size $qN \times qN$ for MV and r blocks of size $N \times N$ for PC model analyses. Similarly, if fitted, genetic groups were held in a single dense block. No distinction between MV and PC was made for this effect, as the transformation yielded sufficient additional coefficients between levels for different traits from the data part, $\mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{Z}^\star$, for the corresponding off-diagonal blocks to be almost dense. The remaining non-zero coefficients in the coefficient matrix were held in compressed sparse row format. A preconditioned conjugate gradient (PCG) algorithm (e.g. Tsuruta *et al.* 2001) with partial Cholesky decomposition preconditioner was used to solve the MME. Cholesky factors and solutions for the dense blocks were obtained using LAPACK routines `DPOTRF` and `DPOTRS` (Anderson *et al.* 1999), respectively. The product of the coefficient matrix and a vector required in each PCG iterate was formed using routines `DSYMV` from the BLAS library (Blackford *et al.* 2002) and the Intel sparse matrix equivalent, `MKL_DCSRSYMV`.

Computations were carried under Linux on a machine with 256GB of RAM and 16 Intel Xeon CPU E5-2630 cores, rated at 2.4Ghz with a cache size of 20MB. BLAS and LAPACK routines used were loaded from the Intel Math Kernel Library (MKL), version 11.1.

Table 1. Computing requirements for equivalent models for 5 traits

			Without genetic groups				With genetic groups			
			Pedigree		Genomic		Pedigree		Genomic	
			MV	PC	MV	PC	MV	PC	MV	PC
No. of equations			12,182,223				12,182,688			
NNZ[a]	Sparse	after data	50.7	66.2	50.7	66.2	223.8	316.0	223.8	316.0
		after random	179.2	89.1	178.9	89.1	352.3	338.9	352.0	338.9
	Dense	genotyped	–	–	1430.6	286.1	–	–	1430.6	286.1
Total			191.4	101.3	1621.6	387.4	364.6	351.2	1794.8	637.2
Memory (GB)			4.3	3.3	25.6	7.8	7.8	7.7	28.8	11.6
No. of PCG iterates			684	693	682	690	1357	1387	1339	1389
Time[b]	single		22.1	19.1	90.5	28.9	44.8	46.4	165.0	64.4
	multi		20.5	20.8	65.3	31.7	42.2	42.9	122.6	61.1

[a]No. of non-zero elements in coefficient matrix (in million) [b]in minutes, for single- and multi-threaded MKL routines

RESULTS

Computational requirements for analyses fitting equivalent models are summarized in Table 1, comparing models with and without the use of genomic information. Values given for NNZ elements pertain to one triangle of the symmetric coefficient matrix. As expected, there were marked differences in the NNZ elements between MV and PC models, with more elements arising from the 'data part' but substantially less non-zero elements due to covariances between random effects for the PC models, especially for single-step analyses. Fitting genetic groups increased the NNZ elements substantially and almost doubled the number of PCG iterates required. PC models proved highly advantageous, with overall computing times reduced 2- to 3-fold when genomic relationships were considered. While CPU time summed over threads when using multi-threaded MKL routines (not shown) seemed to indicate pronounced parallel processing, differences in elapsed time to single-thread runs were surprisingly small, suggesting 'processor spin' rather than actual simultaneous execution.

Corresponding results for a 10-trait scenario, obtained by doubling the data, for single-step models with genetic groups are given in Table 2. Considering more traits amplified differences between models and improved multi-thread performance, especially for the Cholesky decomposition of the diagonal block(s) for genotyped animals in the preconditioning step. Reducing the number of principal components fitted decreased the number of equations in the model and NNZ elements in the coefficient matrix. Results clearly illustrate the increasing advantage of PC over MV models with the number of traits and number of negligible eigenvalues in the genetic covariance matrix among traits.

DISCUSSION

We have described a simple reparameterisation of the standard multivariate mixed model – estimating genetic effects for principal components rather than the traits of interest – and illustrated its potential to reduce computational requirements, especially when parts of the inverse of the relationship matrix are dense. In addition, this parameterisation directly lends itself to dimension reduction by eliminating the principal components which explain virtually no genetic variation, which becomes increasingly important with the number of traits considered. Even a relatively small reduction in dimension can have a big impact on computational requirements with negligible effects on the accuracy of genetic evaluation, if chosen judiciously. Calculations shown for the small subset of traits in LAMBPLAN considered here held the MME in core. In practice, this is unlikely to be feasible and an 'iteration on data' type strategy needs to be employed instead (Tier and Graser 1991). However, the

Table 2. Computing requirements for full and reduced rank models for 10 traits

			MV10	PC10	PC9	PC8	PC7	PC6
No. of equations (in million)			24.37	24.37	22.37	20.37	18.38	16.38
NNZ[a]	Sparse	after groups	901.8	1355.4	1208.4	1015.2	840.3	683.6
		after random	1420.1	1401.2	1249.6	1051.9	872.4	711.1
	Dense	genotyped	5722.4	572.3	515.1	457.8	400.6	343.4
		groups	0.313	0.331	0.271	0.215	0.164	0.121
	Total		7167.1	1989.1	1787.3	1530.2	1291.4	1070.9
Memory (GB)			104.0	28.9	26.4	23.1	20.2	17.3
No. of PCG iterates			1797	1938	1969	1913	1891	1517
Time[b]	single	Precondition	293.3	3.25	3.00	2.57	2.40	1.95
		Total	959	202	188	155	139	126
	multi	Precondition	25.0	0.6	0.5	0.4	0.4	0.3
		Total	551	147	140	127	116	88

[a]No. of non-zero elements in coefficient matrix (in million) [b]in minutes, for single- and multi-threaded MKL routines

NNZ in the coefficient matrix is likely to be at least equally important in such schemes. In addition, if sufficient memory is available, they are readily combined with in-core storage of dense blocks and experience gained here with library routines for matrix computations should be directly transferable.

CONCLUSIONS

Computational strategies described are expected to play an essential rôle in making multi-trait, single-step genetic evaluation for Australian livestock computationally feasible.

ACKNOWLEDGEMENTS

Work was supported by Meat and Livestock Australia grants B.BFG.0050, B.SGN.0027 and B.SGN.0028.

REFERENCES

Aguilar I., Misztal I., Legarra A. and Tsuruta S. (2011) *J. Anim. Breed. Genet.* **128**:422.

Anderson E., Bai Z., Bischof C., Blackford S., Demmel J., Dongarra J., Du Croz J., Greenbaum A., Hammarling S., McKenney A. and Sorensen D. (1999) *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, Third edition.

Blackford L., Demmel J., Dongarra J., Duff I., Hammarling S., Henry G., Heroux M., Kaufman L., Limsdaine A., Petitet A., Pozo R., Remington K. and Whaley R.C. (2002) *ACM Trans. Math. Softw.* **28**:135.

Brown D.J., Huisman A.E., Swan A.A., Graser H.U., Woolaston R.R., Ball A.J., Atkins K.D. and Banks R.G.a. (2007) *Proc. Ass. Advan. Anim. Breed. Genet.* **17**:187.

Meyer K. (2009) *Proc. Ass. Advan. Anim. Breed. Genet.* **18**:442.

Meyer K. and Kirkpatrick M. (2005) *Genet. Sel. Evol.* **37**:1.

Misztal I., Legarra A. and Aguilar I. (2009) *J. Dairy Sci.* **92**:4648.

Misztal I., Vitezica Z.G., Legarra A., Aguilar I. and Swan A.A. (2013) *J. Anim. Breed. Genet.* **130**:252.

Tier B. and Graser H.U. (1991) *J. Anim. Breed. Genet.* **108**:81.

Tsuruta S., Misztal I. and Strandén I. (2001) *J. Anim. Sci.* **79**:1166.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E. and Visscher P.M. (2010) *Nature Genet.* **42**:565.

Proc. Assoc. Advmt. Breed. Genet. 21: 433-436

USING RANDOM FORESTS TO IDENTIFY SNP ASSOCIATED WITH LEG DEFECT IN BROILER CHICKEN: IMPACT OF CORRECTING FOR POPULATION STRUCTURES

Y. Li[1], A. George[2], R. Hawken[3], R. Sapp[3], S. Lehnert[1] and A. Reverter[1]

[1]CSIRO Agriculture Flagship, St Lucia, QLD 4067, Australia
[2]CSIRO Digital Productivity and Services Flagship, Dutton Park, QLD 4102, Australia
[3]Cobb-Vantress Inc., Siloam Springs, Arkansas 72761-1030, USA

SUMMARY

The machine learning method, Random Forests (RF) has been shown to be effective in genome-wide association studies (GWAS). However, the presence of population structure (PS), e.g. relatedness between individuals, may cause spurious results in a RF analysis. In this study, we examined the impact of correcting for PS on the RF analysis of leg defect data from a commercial poultry population of 826 chickens genotyped for 44,129 SNP (single nucleotide polymorphism) markers. The results show that correcting for PS led to: 1) a significant improvement in the estimates of SNP variable importance values; 2) a significant reduction in false positives identified in the uncorrected data; and 3) a stronger evidence for a set of SNPs associated with the defective phenotype.

INTRODUCTION

One of the challenges of GWAS is that the number of predictors is larger than the number of samples, the so called "large p, small n" problem. During the past decades, a number of parametric statistical approaches have been developed for dealing with this issue, for example: Least Absolute Shrinkage and Selection Operator (LASSO) (Wu *et al.* 2009) and two-step Bayesian variable selection method (Zhang *et al.* 2008). Recently non-parametric machine learning methods have been shown to be efficient in analysing large genomic data (Szymazak *et al*. 2009). One of these methods is Random Forests (RF, Breiman 2001; Chen and Ishwaran 2012), a nonparametric decision tree based ensemble method for classification or regression of multiple predictor variables. Our initial preliminary examination found that this method is a powerful tool in pre-screening candidate genes in GWAS of sheep and cattle datasets (Li *et al*. 2014). Despite the advantage of RF over single marker GWAS methods in accounting for correlations among SNP variables, the existence of population structure (PS) has been shown to cause spurious results in the RF analyses (Zhao *et al.* 2012). In this study, we used a dataset from a commercial poultry population to examine the impact of correcting for PS on the RF analysis of a binary trait – leg defect.

MATERIAL AND METHODS

Data. A total of 826 broiler males from a commercial line of Cobb-Vantress Inc. was genotyped for 51,713 SNPs. The dataset comprised animals from 22 generations with various proportion of animals that had leg related problems, ranging from 29% to 51%. After quality check 7,584 SNPs were removed from the genotype dataset and the remaining 44,129 SNPs were used for the RF analyses. The original recording of an animal's phenotypic leg status was either normal, bowed out, bowed in or rotated. We generated a new binary trait, by merging the latter three categories into a single category "Leg Defect'. Of the 826 animals, 592 were normal (coded "0") and 234 had leg defects (coded "1") (Table 1).

EIGENSTRAT analysis for extracting population structure (PS) information. Unlike a linear model that can accommodate PS by fitting a covariance matrix in the model based on pedigree or genomic relationships, RF as a permutation-based method cannot directly account for such factors. Therefore, prior to a RF analysis, it is necessary to identify and correct any existing

population stratification. In this study we applied a method similar to that used by Zhao *et al.* (2012) to correct for PS. An EIGENSTRAT analysis (Price *et al.* 2006) was initially conducted to extract all eigenvectors from the SNP data. The linear regression models were fitted to regress the first 10 axes of variation (principal components) on: a) individual SNP genotypes, and b) the phenotypic trait values, respectively. The residuals from these analyses were combined for the RF analyses. All analyses were performed using the R program (version 3.1.1, http://www.r-project.org/).

Table 1. Trait distribution of leg related defect attributes in 826 roosters.

Trait	Number	0 (Normal)	1 (Defect)
Bowed Out	826	729	97
Bowed In	826	786	40
Rotated Leg	826	729	97
Leg Defect	826	592	234

Random forests (RF). Details of the RF methodology can be found in Breiman (2001). In brief, six steps are involved: 1) As the training dataset, select a random subsample of 550 individuals (or two thirds) with replacement from the available 826 individuals; 2) Select a random subset of SNPs (parameter mtry; say 420 out of the original 44,129) to form a decision tree; 3) Create a single tree via partitioning of sampled individuals in the subsample (normal *versus* defect) with SNP genotypes (e.g. "AA" *versus* others); and with the order (or arrangement) of SNP in the tree run repeatedly until individuals are perfectly partitioned into normal and defect; 4) Test the tree created in Step 3 with the remaining 276 individuals (i.e. validation) to determine the prediction error rate of the SNP tree; 5) Repeat Steps 1 to 4 to develop a large number of forest trees (parameter Ntree); 6) Compute SNP variable importance value (VIM) by averaging the prediction error values across all forest trees. For a continuous phenotype (e.g. corrected data), Step 3 will build a tree that splits the sampled individuals into subsamples with different data value ranges. Step 4 will calculate the minimized sum of squared error for each SNP. It is worth noting that in a RF analysis, a SNP prediction error value is estimated when the SNP is randomly permuted, i.e. excluded from the forest trees. Therefore, the higher the VIM value, the more important the SNP is.

Two crucial parameters impact the outcome of a RF analysis, i.e., the size of forest trees (Ntree) and the number of markers at each sampling event (mtry). To determine the minimum requirement for these parameters, we examined a range of Ntree and mtry values. These included Ntree = 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, or 2000, and mtry = 1, $\sqrt{p}$, $2 * \sqrt{p}$ or $0.1*p$, where p is total number of SNP markers (44,129). Once the minimum parameters were determined, these values were used to run the final RF analyses comprised of 100 RF replicates. To demonstrate the effect of correcting PS on the analysis, we compared the RF results from the corrected data with the uncorrected data. The R program randomForest was used (version 3.1.1).

RESULTS AND DISCUSSION

RF parameter determination. The average SNP VIM values for different parameter combinations of Ntree and mtry are shown in Figure 1. Note that in the context of RF analyses, a high value for VIM is favourable. For both uncorrected and corrected data, the average VIM reached a stable status with Ntree ≥ 1,000. This suggests that the RF analysis with Ntree ≥ 1,000 should produce reasonably accurate VIM values. Among the four parameters tested for mtry, single marker analysis (mtry =1) gave the lowest estimates for VIM, while the other three parameters ($\sqrt{p}$, $2 * \sqrt{p}$ and $0.1*p$) produced very similar values.

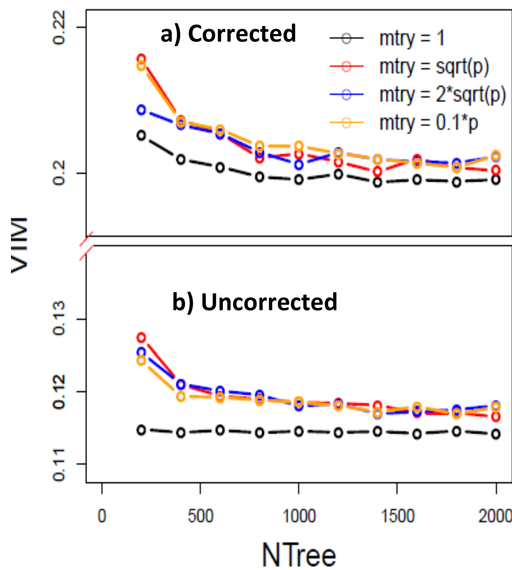


Figure 1. Comparison of the mean VIM values from different combinations of parameters with (top panel) and without (bottom panel) correction for population structure.

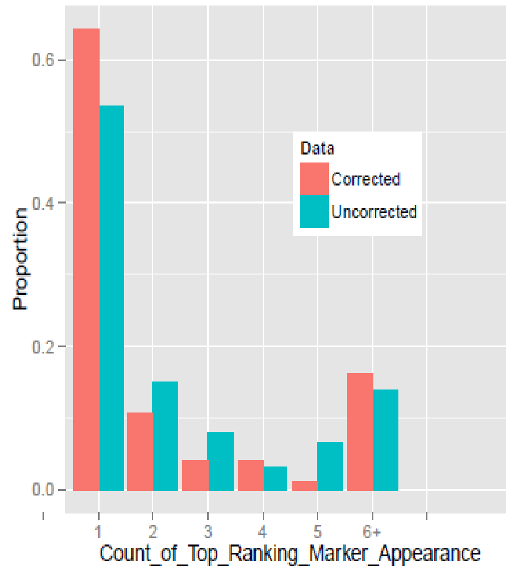


Figure 2. The proportions of top 5 marker appearances in 100 RF replicates with (red bars) and without (blue bars) correction for population structure.

RF analyses in the corrected and uncorrected datasets. Compared to the uncorrected data (Figure 1b), correcting for PS (Figure 1a) clearly resulted in a significant increment in the estimated average VIM values (from 0.120 to 0.205). When investigating the top 5 ranking markers from each of the 100 RF replicates, a total of 166 and 179 markers were found in the uncorrected and corrected data, respectively. The compositions of these marker incidences in both datasets are shown in Figure 2. The top markers appearing only once in 100 replicates had the highest proportion (54% in the uncorrected versus 64% in the corrected data). The uncorrected data tended to have fewer markers (13.85%) with the highest incidence (i.e. captured in 6+ replicates) than the corrected data (16.20%). However, the intriguing results were found when comparing the distributions of the top 5 marker incidences across the whole genome in both datasets (Figure 3). It is very clear that correcting for PS led to a reduction in top ranking SNP incidence in a number of genome regions found to be significant in the leg defect analysis of the uncorrected dataset. The majority SNPs identified in the uncorrected data were no longer in the top ranking markers in the corrected data. Among 166 (uncorrected data) and 179 (corrected data) markers, there were 26 in common (shown by overlapping regions in Figure 3) and 11 of them had a reduced incidence in the corrected data. In contrast, there was a set of 12 common markers closely linked (near the right hand side of the genome), after correcting for PS, the association signal became much stronger.

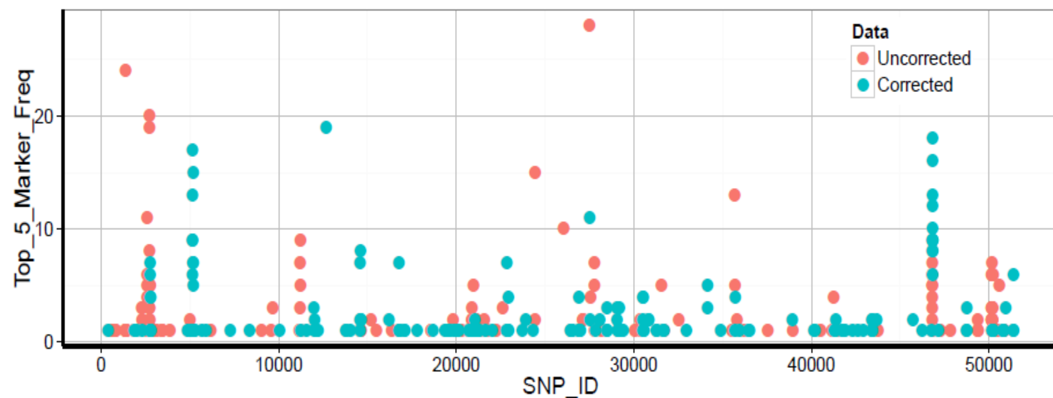


Figure 3. Distribution of top 5 ranking marker incidences across genome for corrected (blue) and uncorrected data (red).

Population stratification or admixture is known to cause different allele or genotype frequencies in subpopulations and that in turn can lead to detection of spurious SNP associations in GWAS (Zhao *et al*. 2012). RF has its advantage over single marker GWAS methods in handling high dimensional genomic data (Chen and Ishwaran 2012), but it has a limited power in dealing with a confounding effect of PS on both genotypes and a phenotype. The results here demonstrate the importance of correcting for population structure prior to RF analysis to minimize false positives. Since the "true" SNPs are unknown, these results are of very limited use for the purpose of method validation. There is a need in future to conduct a systematic evaluation of the method with large simulation datasets.

CONCLUSIONS

Correcting for population structure prior to RF analysis can improve the accuracy of SNP variable importance values and avoid spurious association results. Since RF is a non-parametric permutation based method, a large number of RF replicates is required to get reliable inference of the markers associated with a phenotype.

REFERENECES

Breiman L. (2001) *Machine Learning*. **45**: 5.
Chen X. and Ishwaran H. (2012) *Genomics*. **99**: 323.
Li Y., Kijas J., Henshall J., Lehnert S.A., McCulloch M. and Reverter A. (2014) 10th WCGALP.
Price A.L, Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. and Reich D. (2006). *Nature Genet.* **38:** 904.
Szymazak S., Biernacka J.M., Cordell H.J., Gonzalez-Recio O., Konig I.R., Zhang H. and Sun Y. (2009) *Genet. Epidemiol.* **33** (Supplement 1): S51.
Wu T.T., Chen Y.F., Hastie T., Sobel E. and Lange K. (2009) *Bioinformatics*. **25**: 714.
Zhang M., Zhang D. and Wells M.T. (2008) *BMC Bioinformatics*. **9**: 251.
Zhao Y., Chen F., Zhai R., Lin X., Wang Z., Su L. and Charistiani D.C. (2012) *Int. J. Epidemiol*. **41**: 1798.

Proc. Assoc. Advmt. Breed. Genet. 21: 437-440

MISSING HERITABILITY OF ADAPTATION PHENOTYPES IN TROPICAL CATTLE

Laercio R. Porto-Neto, Nicholas J. Hudson, William Barendse, Sigrid A. Lehnert and Antonio Reverter

CSIRO Agriculture Flagship, 306 Carmody Road, St Lucia, QLD 4067, Australia

SUMMARY

The ‘missing heritability problem’ is the inability to tag all the genetic variance of a trait using genome wide single nucleotide polymorphism. Here, we compute missing heritability for two populations of cattle phenotyped for ten tropical adaptation traits that exhibit variable genetic architectures. We derived genomic relationship matrices (GRM) using both low and high density SNP panels, and computed the missing heritability through comparison to pedigree (NRM). Overall, the low density indicine panel performs very well in characterising the Brahman population. We found that estimation of missing heritability was broadly similar for both panels across the ten phenotypes. This implies similar amounts of genetic variation relevant to those phenotypes have been captured. The phenotypes with the lowest missing heritability (coat type and sheath score in Tropical Composites) possess an architecture that can be characterised simply. That is, they are dominated by genes of large effect.

INTRODUCTION

The total genetic variance of a trait is usually estimated using pedigree information. Then the total variance (phenotypic variance) is partitioned into genetic and environmental variance. The heritability or ratio of genetic variance to total phenotypic variance can be calculated. In genome-wide association studies (GWAS), the shrunk variance associated with each significant single nucleotide polymorphism (SNP) can be estimated. If all SNP are considered simultaneously, a large proportion of the variance that would have been missed due to small individual effects can be captured (Yang *et al*. 2010). However, it has been shown that in all studies to date this sum of those variances is usually far less than the total genetic variance. Since the SNP in the GWAS cover 90% or more of the whole genome, the inability to account for the total genetic variance is called the missing heritability problem.

The degree of missing heritability varies for different traits in the same dataset. Here we estimate the different degrees of missing heritability for 10 traits of tropical cattle for two breeds to determine whether there are large differences either between traits or between breeds in the amount of the genetic variance that is not tagged by SNP.

To estimate the missing heritability we use the approach of simultaneously analysing the genetic variance using both pedigree and SNP data, in which the genetic variance not accounted for by the SNPs will be assigned to the pedigree component. This allows us to maximize the amount of genetic variance assigned to SNP, avoids the issues of significance thresholds, and biased estimation of the SNP effects due to small sample sizes (in the thousands instead of in the hundreds of thousands).

MATERIAL AND METHODS

Animals, genotypes and phenotypes. We used the genetic and phenotypic resources outlined in Porto-Neto *et al.* 2014. In brief, 2112 Brahman and 2550 Tropical Composite animals were genotyped with either the Bovine HD BeadChip (Illumina Inc., San Diego, CA) that includes more than 770,962 SNP or the GGP Indicine chip that includes 71,726 SNP.

Calculation of genetic relationship matrix (GRM) and numerator relationship matrix (NRM). The GRM was computed based on the methodology developed by Van Raden (2008).

$$GRM = \frac{\mathbf{ZZ}^T}{2\sum p_i(1-p_i)}$$

where **Z** is a matrix that relates SNP alleles to individuals and p_i is the frequency of the second allele for the i-th SNP. $\mathbf{ZZ}^\mathrm{T}$ represents the number of shared SNP alleles among two individuals and the division of $\mathbf{ZZ}^\mathrm{T}$ by $2\sum p_i(1-p_i)$ aims at scaling the GRM to make it analogous to the NRM obtained based on the pedigree information.

Variance components. A single model with two random components was run in Qxpak (Perez-Enciso and Misztal 2011). We estimated missing heritability using the approach of Roman-Ponce *et al*. 2014.

$$C_{miss} = 1 - \frac{\sigma_u^2}{\sigma_a^2 + \sigma_u^2}$$

where σ_u^2 is the variance due to the genotype data (ie GRM) and σ_a^2 is the additive genetic variance due to the pedigree (ie. the NRM in our context).

RESULTS AND DISCUSSION

To provide an estimate of how well the two SNP panels characterised the genetics of the two populations, we plotted the population level frequency of the reference allele of each loci on a genome-wide basis (Figure 1).

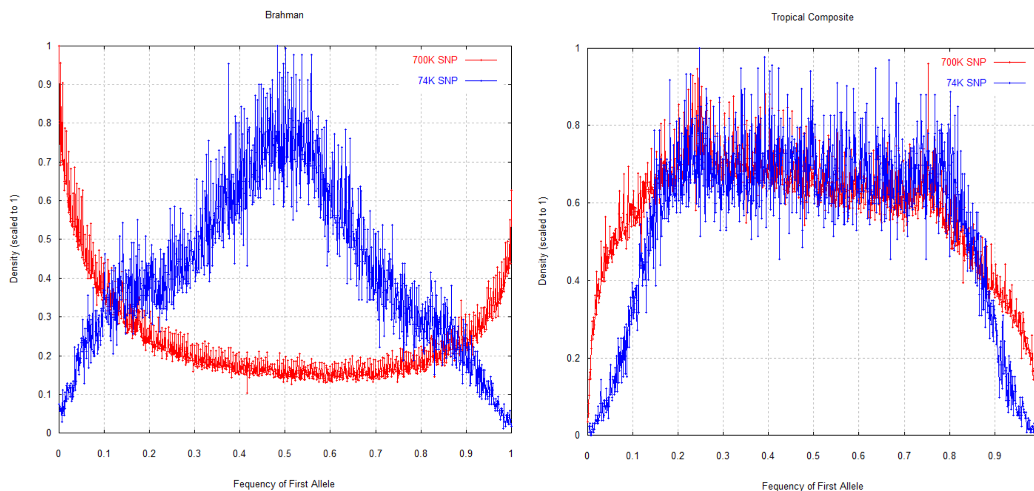


Figure 1. Population level allelic frequency for each SNP panel, 74K (blue lines) and 700K (red lines) on the Brahman (left hand plot) and Tropical Composite (right hand plot) populations.

The left hand plot shows that the Brahman sample has a relative deficiency of low frequency alleles in the low density Indicine SNP panel (74K), compared to the HD SNP (700K) panel, and the frequency of the reference allele is biased to p = 0.5. On the other hand, there is a bias on the HD SNP panel towards low frequent alleles. The right hand plot shows that the Tropical Composite animals have an allele frequency distribution that is similar for both the low density

Indicine SNP panel as well as the HD SNP panel, although there does seem to be a relative lack of very low frequency SNPs in the low density panel for this sample. It is worth mentioning that the 74K selection is part of the full HD panel, so the impact of two independent SNP selections was not tested here. We could speculate that a set of SNP that poorly characterise the structure of a population would also poorly perform in capturing the traits' variance. Put another way, a random selection of 74K SNP would perform worst than a set of SNP with on average high minor allelic frequencies.

We next computed missing heritability scores for the Brahman (Table 1) and Tropical Composite (Table 2) using both SNP panels.

Table 1. Genetic parameters and missing heritability for Brahman cattle based on low and high density SNP panels.

TRAIT	74K			700K		
	NRM	GRM	Missing	NRM	GRM	Missing
Coat type	0.232	0.264	0.468	0.228	0.269	0.460
Coat colour	0.269	0.324	0.453	0.186	0.362	0.339
Condition score	0.223	0.375	0.372	0.139	0.409	0.253
Worm eggs (n/gr)	0.351	0.363	0.491	0.351	0.365	0.491
Fly lesions	0.231	0.282	0.450	0.261	0.287	0.476
Flight time	0.286	0.233	0.552	0.287	0.233	0.552
Sheath score	0.284	0.335	0.459	0.281	0.360	0.438
Temperature	0.230	0.195	0.541	0.233	0.187	0.555
Tick score	0.413	0.384	0.518	0.413	0.383	0.519
Yearling Weight	0.210	0.316	0.399	0.208	0.306	0.405

Table 2. Genetic parameters and missing heritability for Tropical Composite cattle based on low and high density SNP panels.

TRAIT	74K			700K		
	NRM	GRM	Missing	NRM	GRM	Missing
Coat type	0.169	0.502	0.252	0.267	0.430	0.383
Coat Colour	0.284	0.389	0.422	0.279	0.400	0.411
Condition score	0.175	0.321	0.354	0.175	0.306	0.365
Worm eggs (n/gr)	0.369	0.359	0.507	0.365	0.365	0.500
Fly lesions	0.359	0.364	0.497	0.358	0.365	0.495
Flight time	0.251	0.311	0.447	0.241	0.327	0.424
Sheath score	0.235	0.529	0.308	0.235	0.531	0.307
Temperature	0.218	0.157	0.581	0.215	0.156	0.581
Tick score	0.382	0.370	0.508	0.387	0.375	0.507
Yearling Weight	0.293	0.363	0.446	0.285	0.385	0.425

These missing heritability estimates are somewhat higher than the 36.6% reported by Roman-Ponce *et al* (2014) across 11 cattle traits and the 38.5% reported by Haile-Mariam *et al* (2013) across 29 cattle traits. In the Brahman sample, the extra genetic information provided by the high density SNP panel only led to a systematic reduction in missing heritabilities in coat colour and condition score. In the Tropical Composite sample, the larger SNP panel did not lead to a clear reduction in missing heritability, and interestingly, in the trait coat type the missing heritability was lower when the lower density panel was used. This particular phenotype is controlled by a small number of genes and so the addition of non-informative SNP from across the genome may have reduced the accuracy of prediction.

We noted that the two phenotypes with the lowest missing heritability in the Tropical Composite populations (Coat type and Sheath score) are phenotypes with a relatively simple genetic architecture characterised by relatively few gene regions of large effect (Porto Neto *et al.* 2014). It is not surprising that this kind of simple genetic architecture would be more amenable to accurate modelling of genotype to phenotype relationships.

REFERENCES

Haile-Mariam M., Nieuwhof G.J., Beard K.T., Konstatinov K.V., Hayes B.J. (2013) *J. Anim. Breed. Genet.* **130**: 20.

Perez-Enciso M. and Misztal I. (2011) *BMC Bioinf.* **12**: 202.

Porto-Neto L.R., Reverter A., Prayaga K.C., Chan E.K.F., Johnston D.J., Hawken R.J., Fordyce G., Garcia J.F., Sonstegard T.S., Bolormaa S., Goddard M.E., Burrow H.M., Henshall J.M., Lehnert S.A. and Barendse W. (2014) *PLoS ONE* **9**: e113284.

Roman-Ponce S.I., Samore A.B., Dolezal M.A., Bagnato A., Meuwissen T.H.E. (2014) *Genet. Selec. Evol.* **46**: 36.

Van Raden P.M. (2008) *J. Dairy Sci* **91**: 4414.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E. & Visscher P.M. (2010) *Nature Genetics* **42**: 565.

Proc. Assoc. Advmt. Breed. Genet. 21: 441-444

IMPUTATION ACCURACY MEASUREMENT AND POST-IMPUTATION QUALITY IN IMPUTED SNP GENOTYPES FOR DAIRY CATTLE

K.M. Tiplady and R.G. Sherlock

LIC, Private Bag 3016, Hamilton 3240, New Zealand

SUMMARY

Imputation of genotypes is a cost-effective method for generating genotypes for un-typed loci and allows data from different genotyping panels and platforms to be combined. Accuracy of imputation can be defined in a number of ways to distinguish well-imputed from poorly-imputed SNP. The aims of this study were to compare different measures of imputation accuracy in low density panel data and determine how well the estimated allelic R^2 (AR^2) measure reported by BEAGLE performs across minor allele frequency (MAF) as a post-imputation filtering tool. Genotypes for 28,793 New Zealand mixed-breed dairy cows from a low density BeadChip (n=16,512 SNP) were used in the study. For 17,593 animals, 9,166 SNP were masked and imputed using version 4.0 of BEAGLE software. Imputation accuracy for SNP with MAF ≥ 0.005 was high, but was variable for low MAF (< 0.005) SNP. Genotypic concordance was not informative for low MAF SNP and was poorly correlated with AR^2 for low MAF SNP. Other imputation accuracy measures (genotypic correlation, minor allele sensitivity and imputation quality score) were informative for low MAF SNP and were highly correlated with AR^2 across all MAF classifications (r > 0.81). Results showed that post-imputation filtering based on AR^2 is an effective approach for removing poorly-imputed SNP, including those of low MAF.

INTRODUCTION

Genotype imputation increases the power of existing data by providing predicted genotypes for loci that have not been directly assayed. It allows data from different genotyping platforms to be combined and makes additional variants available for analysis without the cost of actually genotyping them. Compared to using a smaller set of only true genotypes, the additional power from imputed genotypes can provide better signal in genome wide association studies (Khatkar *et al.* 2013) and better estimates of direct genetic values (Khatkar *et al.* 2012; Weigel *et al.* 2010). However, incorrectly imputed genotypes can add noise and compromise an analysis (Weigel *et al.* 2010; Chen *et al.* 2014). Imputation correctness has been evaluated based on a number of accuracy metrics in previous studies (Khatkar *et al.* 2013; Calus *et al.* 2014), each providing a different way to distinguish well-imputed from poorly-imputed SNP. This differentiation can be particularly problematic for low minor allele frequency (MAF) SNP where accurate imputation is more difficult and sensitive to genotype calling errors (Lin *et al.* 2010; Calus *et al.* 2014). Also, some measures of accuracy are highly dependent on MAF and can give misleading results for low MAF SNP (Lin *et al.* 2010; Hickey *et al.* 2012). In this study, accuracy of imputation was examined for genotypes from New Zealand (NZ) progeny test dairy herds which were genotyped on a custom GGP-LD BeadChip. Imputing low MAF SNP well is important within this context because these custom SNP chip panels are often updated with new loci, many of which are low MAF, with a requirement for these to be imputed through the historically genotyped population. Generating imputation accuracy metrics requires a comparison set of true and imputed genotypes, and this is often obtained by selecting a subset of animals as a validation set. For this validation subset, a set of SNP of interest are masked and then imputed. In practical applications of imputation where a complete "truth set" is unavailable, pedigree relationships can sometimes be used to infer true genotypes and the level of imputation accuracy. However, a generally-available post-imputation quality measure which is not dependent on having a "truth set" and is reliable across MAF is

desirable. Browning and Browning (2009) outline a post-imputation estimate of imputation accuracy, the estimated allelic R^2 (AR^2) which is not dependent on allele frequency or having a "truth set" of genotypes. The aims of this study were to compare the AR^2 reported by BEAGLE (Browning and Browning 2009) to a number of different imputation accuracy metrics derived from comparing true with imputed genotypes, and determine how well the AR^2 performs across MAF as a post-imputation filtering tool.

MATERIALS AND METHODS

Genotypes from New Zealand (NZ) progeny test dairy herds (Holstein-Friesian, Jersey and crossbreed) were obtained from a custom version of the GGP-LD BeadChip with 20,183 SNP. After removing animals with a call rate < 0.95 and any SNP that were non-autosomal or had a call rate < 0.9, 19,143 SNP for each of 28,793 animals were included in the study.

Imputation reference. Reference animals were selected as those with progeny in the wider population (11,062 females; 138 males). Average pedigree relationships between reference animals were 0.034 (sd=0.031). Monomorphic SNP were removed and missing SNP were imputed using version 4.0 of BEAGLE (Browning and Browning 2009) with default parameters. This resulted in an imputation reference of 16,512 SNP for 11,200 animals.

Imputation target. Genotypes for 17,593 animals not included in the imputation reference were included in the imputation target population. Of the target population, 38.4% had at least 1 parent in the reference, and the average pedigree relationship between reference and target animals was 0.033 (sd=0.029). Of the 16,512 SNP in the imputation reference, 9,166 were masked to leave only the SNP in common with an earlier version of the GGP-LD BeadChip. Imputation was carried out using version 4.0 of BEAGLE with default parameters. True and imputed genotypes were compared for 9,166 masked SNP on 17,593 animals.

Imputation accuracy. Imputation accuracy was assessed according to 4 measures: Genotypic concordance (GCONC; proportion of genotype calls where the true genotype matches the most likely imputed genotype), genotypic correlation (GCORR; correlation between observed and imputed number of copies of the alternate allele), minor allele sensitivity (MAS; proportion of times a minor allele is correctly called when it is present, analogous to non-reference sensitivity) and imputation quality score (IQS; concordance adjusted for chance agreement) as defined by Lin *et al.* (2010).

Post-imputation quality. Post-imputation quality was assessed using the AR^2 calculated by BEAGLE. This is an estimate of the squared correlation between the allele dosage of the most likely imputed genotype and the allele dosage of the true genotype. The true genotype is unknown but the allelic R^2 is estimated from the distribution of imputed posterior genotype probabilities.

MAF classifications. SNP were grouped by frequency of the minor allele in the reference.

RESULTS AND DISCUSSION

Table 1 summarises imputation accuracy as measured by GCONC, GCORR, MAS and IQS, and the AR^2 reported by BEAGLE. For SNP with MAF < 0.005, GCORR, MAS and IQS all indicated measures of accuracy ≤ 0.462, whereas GCONC indicated a high accuracy (0.999). Also, a decrease in GCONC was observed with increasing MAF, but an increase in accuracy was observed when measured by GCORR, MAS and IQS. This is because GCONC is dependent on MAF, and demonstrates that measuring accuracy based on GCONC can be misleading for low MAF SNP, as outlined by Calus *et al.* (2014). Mean AR^2 values also increased with MAF and were particularly low (0.188) for SNP with MAF < 0.005. Imputation accuracy levels were high (≥ 0.864) when MAF ≥ 0.005 based on all 4 measures considered in this study. The MAF at which SNP are accurately imputed would be expected to increase as the size of the imputation reference decreases.

Table 1. Mean imputation accuracy (GCONC, GCORR, MAS, IQS) and post-imputation quality (AR^2) for SNP classified by MAF.

MAF classification	N	GCONC	GCORR	MAS	IQS	AR^2
< 0.005	1218	0.999	0.462	0.198	0.217	0.188
0.005-0.01	130	0.998	0.906	0.864	0.887	0.774
0.01-0.05	557	0.995	0.951	0.930	0.947	0.873
≥ 0.05	7261	0.974	0.965	0.974	0.951	0.910
All	9166	0.979	0.928	0.867	0.852	0.810

Correlations between imputation accuracy measures and AR^2 are shown in Table 2. GCONC was poorly correlated with AR^2 for SNP with MAF < 0.005. Other imputation accuracy measures (GCORR, MAS, IQS) were highly correlated (≥ 0.812) with AR^2 across all minor allele frequencies. High correlations between these accuracy measures and AR^2 suggest that AR^2 may be a good tool for screening SNP post-imputation.

Table 2. Correlations between AR^2 and imputation accuracy (GCONC, GCORR, MAS, IQS) classified by MAF.

MAF classification	N	GCONC	GCORR	MAS	IQS
< 0.005	1218	-0.069	0.851	0.908	0.903
0.005-0.01	130	0.763	0.824	0.852	0.900
0.01-0.05	557	0.644	0.837	0.846	0.882
≥ 0.05	7261	0.888	0.925	0.812	0.947
All	9166	-0.077	0.927	0.972	0.974

Figure 1 shows the distribution of GCORR values prior to and post filtering based on an AR^2 threshold of 0.7. Prior to filtering imputed genotypes, GCORR values were highly variable, in particular for SNPs with MAF < 0.005 (Figure 1a). After filtering, the variation in GCORR values was significantly reduced, particularly for SNP with MAF < 0.005 (Figure 1b). In total, 1191 SNP were removed, most of which were SNP with MAF < 0.005. Results for MAS and IQS were similar (not presented here).

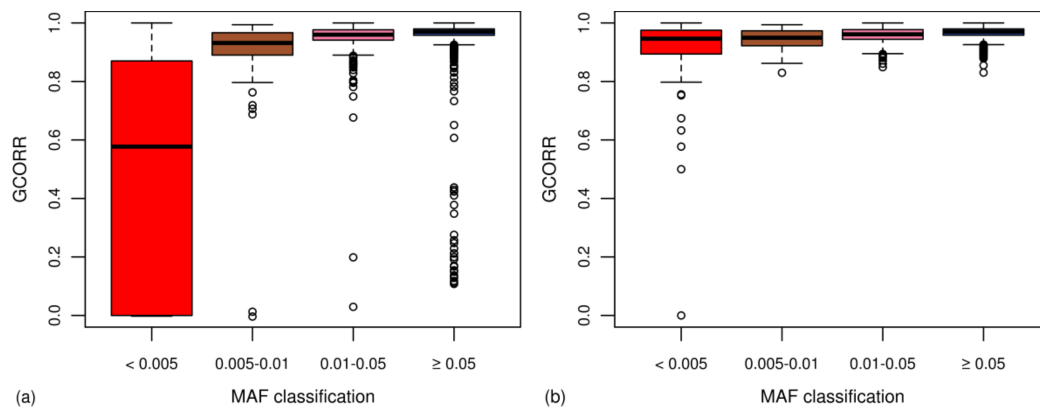


Figure 1. Distribution of genotypic correlation (GCORR) (a) prior to filtering and (b) post filtering based on an AR^2 threshold of 0.7.

Browning and Browning (2009) demonstrated that at high SNP density, AR^2 is a good metric for estimating imputation accuracy without dependence on allele frequency. Kelly *et al.* (2013) also showed that in a population of composite tropical cattle, AR^2 was an effective measure for identifying a large number of poorly-imputed SNP when imputing from Illumina BovineSNP50 to Illumina BovineHD SNP panels. Table 3 summarises mis-classifications of SNP in this study that resulted when a post-imputation filter of $AR^2 > 0.7$ was used to predict SNP that had been imputed well according to each of the accuracy measures GCORR, MAS and IQS. For each measure, well-imputed SNP are defined as those where the measure was > 0.7. False negative (FN) SNP were defined as those with an $AR^2 \leq 0.7$ but an imputation accuracy > 0.7. False positive (FP) SNP were defined as those with an $AR^2 > 0.7$ but an imputation accuracy ≤ 0.7. Low FN rates (≤ 3.77%) were observed for SNP with MAF ≥ 0.01, but were higher for SNP with MAF < 0.01 (5.83-22.88%). Very low FP rates (≤ 0.9%) were observed for SNP with MAF < 0.005 and were all zero for SNP with MAF ≥ 0.005. These results confirm that post-imputation filtering based on AR^2 is an effective approach for removing poorly-imputed SNP, including those of low MAF.

Table 3. Percentage of false positive (FP) and false negative (FN) SNP for imputation accuracy measures (GCORR, MAS, IQS) based on an AR^2 threshold of 0.7.

	GCORR		MAS		IQS	
MAF classification	FN	FP	FN	FP	FN	FP
< 0.005	22.88	0.85	5.83	0.90	9.52	0.41
0.005-0.01	20.93	0	17.69	0	16.15	0
0.01-0.05	3.77	0	3.41	0	3.41	0
≥ 0.05	0.25	0	0.36	0	0.15	0
All	2.35	0.06	1.52	0.12	1.82	0.05

CONCLUSION

Genotypic concordance was not informative for low MAF SNP and was poorly correlated with AR^2 values reported by BEAGLE for low MAF SNP. Other imputation accuracy measures examined (GCORR, MAS, IQS) were informative for SNP across all minor allele frequencies. These measures were also highly correlated with each other and with post-imputation AR^2 values. Post-imputation filtering based on an AR^2 threshold of 0.7 was shown to be an effective way of removing poorly-imputed SNP for imputed genotypes from a population of NZ dairy cows genotyped on a low density panel.

REFERENCES

Browning B.L. and Browning S.R. (2009) *Am J Hum Genet* **84**: 210.

Calus M.P., Bouwman A.C., Hickey J.M., Veerkamp R.F. and Mulder H.A. (2014) *Animal* **8**: 1743.

Chen L., Li C., Sargolzaei M. and Schenkel F. (2014) *PLoS ONE*, **9(7)**: e101544.

Hickey J.M., Crossa J., Babu R. and de los Campos G. (2012) *Crop Science* **52:** 654.

Kelly M., Fortes M.R.S. and Moore S.S. (2013) *Proc. Assoc. Advmt. Anim. Breed Genet.* **20**: 550-553.

Khatkar M.S., Moser G., Hayes B.J. and Raadsma H.W. (2012) *BMC Genomics* **13**: 538.

Khatkar M.S., Thomson P.C. and Raadsma H.W. (2013) *Proc. Assoc. Advmt. Anim. Breed Genet.* **20**: 554-557.

Lin P., Hartz S. M., Zhang Z., Saccone S.F., Wang J., *et al.* (2010) *PLoS ONE* **5(3)**: e9697.

Weigel K.A., de Los Campos G. and Vazquez A., *et al.* (2010) *J. Dairy Sci.* **93**: 5423.

Proc. Assoc. Advmt. Breed. Genet. 21: 445-448

INBREEDING DEPRESSION IN ADAPTATION PHENOTYPES OF TROPICAL BEEF CATTLE USING GENOME WIDE DATA

L.R. Porto-Neto[1], N.J. Hudson[1], W. Barendse[1], M.A.R. de Cara[2], S.A. Lehnert[1] and A. Reverter[1]

[1] CSIRO Agriculture Flagship, Queensland Bioscience Precinct, Brisbane, Queensland, Australia
[2] Laboratoire d'Eco-anthropologie et Ethnobiologie, UMR 7206 MNHN, UPD7, Paris, France

SUMMARY

Inbreeding has the potential to negatively impact on animal performance. Strategies to monitor and mitigate inbreeding depression require that it can be estimated accurately. Here, we use genome-wide SNP data to produce three alternative estimates of inbreeding: proportion of heterozygous SNP, diagonal elements of the genomic relationship matrix and runs of homozygosity. We focus on Brahman (N = 2,112) *versus* Tropical Composite (N = 2,550) cattle which have been genotyped by a 74K Indicine chip and phenotyped for 10 traits of commercial importance to tropical adaptation. While the summary statistics for the Brahman and Tropical Composite populations are remarkably similar, the relationships of the inbreeding coefficients to the various phenotypes are different. In both populations, the proportion of heterozygous SNP was positively correlated with body condition score and negatively correlated with yearling weight. A possible explanation relating to the content of the Indicine SNP chip is briefly discussed.

INTRODUCTION

Inbreeding can reduce animal performance for commercially important traits. In order to mitigate inbreeding through management strategies and breeding programs, it is important to be able to estimate it accurately. Traditionally, inbreeding coefficients (F) have been computed from pedigree information. Given the recent availability of genome-wide single nucleotide polymorphism (SNP) data, it is now possible to exploit real genetic data to infer F. According to Saura *et al*. (2015) genomic F has the following three advantages: it measures homozygosity directly rather than its expectation; it can be applied to particular genomic regions such as those harbouring QTL; and it can be estimated with incomplete or even absent pedigree information.

Here, we explore three different measures of inbreeding using 71,726 SNP genotypes from Brahman (BB) and Tropical Composite (TC) cattle measured for 10 phenotypes of relevance to tropical adaptation. The three inbreeding measures we used are: proportion of heterozygous SNP (HET_F), diagonal elements of the genomic relationship matrix (GRM_F) and runs of homozygosity (ROH_F) (Keller *et al.* 2011; Saura *et al.* 2015). Sample size is relatively large (2,112 Brahman and 2,550 Tropical Composite) and genome-wide association studies have already been well characterised (Porto-Neto *et al.* 2014).

MATERIALS AND METHODS

Animals, genotypes and phenotypes. We used genetic and phenotypic resources outlined in Porto-Neto *et al.* (2014). In brief, 2,112 BB and 2,550 TC cattle with genotypes for 729,068 SNP. For the present work we extracted the 71,726 SNP corresponding to the GGP Indicus HD chip (http://www.neogeneurope.com/Genomics/pdf/Slicks/NE_GeneSeekCustomChipFlyer.pdf). The ten phenotypes are: coat type (COAT), coat colour (COLOR), condition score (COND), worm eggs (EPG), fly lesions (FLY), flight time (FT), navel/sheath score (NAVEL), temperature (TEMP), tick score (TICK) and yearling weight (YWT). In addition, we used the SNP-based estimated zebu content also described in Porto-Neto *et al.* (2014) as an additional variable to be related with the measures of inbreeding described next.

Measures of inbreeding. We explored three alternative measures of inbreeding:

(1) GRM_F: Genomic inbreeding based on the diagonal elements of the genomic relationship matrix (GRM). The GRM was computed according to Van Raden (2008) and allele frequencies calculated separately for the Brahman and Tropical Composite population;
(2) HET_F: Proportion of the total SNP genotypes that were heterozygous;
(3) ROH_F: Proportion of the genome that consists of runs of homozygosity (ROH).

We followed computational approaches described in Saura *et al*. (2015) by which ROH was detected according to the following criteria: (1) At least 10 kb in a ROH window; (2) At most one heterozygous SNP; (3) A density of at least 1 SNP per 100 kb; (4) A maximum distance of 100 kb between two SNP in a given ROH.

Inbreeding depression. The effect of each measure of inbreeding on the phenotypes was estimated by fitting a model that contain the phenotype as dependent variable and the measures of inbreeding plus the zebu content as fixed regression covariates as well as the fixed class effects of contemporary group (combination of sex, year and location), age of dam. All analyses were performed using SAS 9.3 (SAS Inst., Cary, NC) one phenotype at a time and separately for the BB and the TC datasets.

RESULTS AND DISCUSSION

Table 1 shows summary statistics for the three measures of inbreeding in the two populations. For a given measure, the means are very similar in the two populations and they are also similar to the equivalent values reported by Pryce *et al.* (2014) with dairy cattle. In that work, the authors reported an average GRM_F of 1.134 and 1.144 for Holstein and Jersey, respectively. They also reported an average HET_F of 0.347 and 0.285 for the same two breeds of dairy cattle.

In spite of the overall similarities between the two populations for the 3 inbreeding measures (Table 1), there are remarkable differences in the way they correlate with the phenotypes and zebu content. A likely confound is the indicine chip which better characterises the genetic variation in the Brahman population than it does in the Tropical Composite. As expected, we find that a low heterozygosity is correlated with inbreeding in Brahman ($r = -0.271$; $P < 0.0001$). However, this relationship is actually reversed for the Tropical Composite cattle ($r = 0.495$; $P < 0.0001$).

Table 1. Summary statistics for the three measures of inbreeding in the two populations.

Population	Measure of Inbreeding	Mean	Standard Deviation	Minimum	Maximum
Brahman	HET_F	0.391	0.018	0.284	0.456
(N = 2,112)	GRM_F	1.021	0.039	0.931	1.691
	ROH_F	0.376	0.020	0.306	0.533
Tropical Composite	HET_F	0.384	0.029	0.234	0.481
(N = 2,550)	GRM_F	1.021	0.067	0.880	1.412
	ROH_F	0.382	0.032	0.278	0.566

In both populations, HET_F was positively correlated with COND and negatively correlated with YWT (Figure 1). The strongest correlation involving ROH_F was with YWT being negative for Brahman (r = -0.175; P < 0.0001) and positive for Tropical Composite (r = 0.211; P < 0.0001). The discrepancies between the populations may reflect spurious results from the Tropical Composite. These animals are only 27.2% Zebu on average (Porto-Neto *et al.* 2014) which may lead to artefacts arising from the reliance on the Indicine chip for assaying gene variants.

An alternative explanation is that these correlations, while statistically significantly different from zero, are rather small in magnitude and hence of little biological relevance, and their significance the result of the relatively large sample sizes. This feature is clearly illustrated by Figure 2 showing the scatter plot between ROH_F and YWT for the two populations.

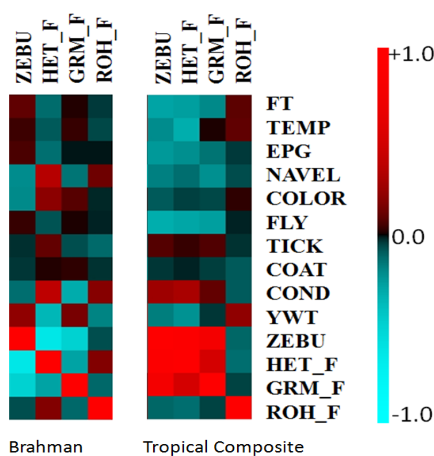


Figure 1. Heat map of the correlation matrix between zebu content and the three measures of inbreeding with each other and the ten phenotypes and for the two populations.

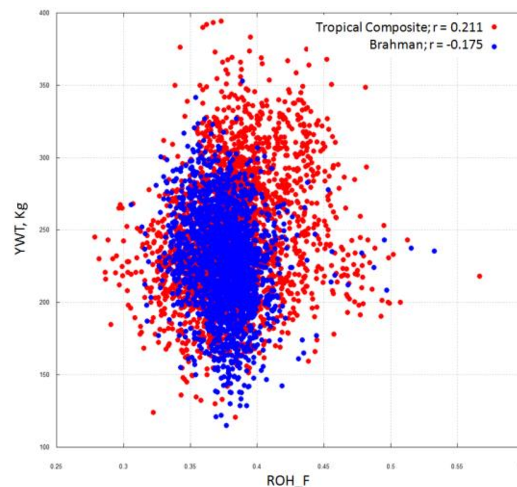


Figure 2. Scatter plot of the relationship between inbreeding based on runs of homozygocity (ROH_F) and yearling weight (YWT) for the Brahman (blue dots) and Tropical Composite (red dots).

In the Brahman dataset, the linear model used for the estimation of inbreeding depression yielded an R^2 that ranged from 18.4% for COLOR to 67.7% for YWT. Similarly, in the Tropical Composite, the R^2 ranged from 18.6% for TICK to 75.6% for YWT. Table 2 shows the estimates of inbreeding depression for each phenotype and in the two populations. In the Brahman population, significant inbreeding depression estimates were identified for FT, EPG, COLOR, COND and YWT. This contrasts with the Tropical Composite population for which only TEMP and YWT was significantly affected by either measure of inbreeding.

While HET_F and ROH_F had very similar mean and variability (Table 1), they were poorly correlated (0.171 and -0.122 in the Brahman and Tropical Composite population, respectively) which may explain the large differences observed for some of the estimates of inbreeding depression. Consistent among both populations was the effect of inbreeding depression on YWT. In the Brahman population, a 1% increase in HET_F was associated with an increase of 1.89 kg of YWT. Similarly, a 1% increase inbreeding measured by GRM_F was associated with a decrease of 1.08 kg of YWT in the Tropical Composite population.

Table 2. Estimates of inbreeding depression for the three measures of inbreeding in the ten phenotypes and two populations.

Phenotype	Brahman			Tropical Composite		
	HET_F	GRM_F	ROH_F	HET_F	GRM_F	ROH_F
FT	-101.44	-73.54	145.69*	-170.20	-14.30	-22.15
TEMP	1.01	0.20	-0.25	-2.64	0.28	1.14*
EPG	-3,097.96**	-995.83**	-329.25	-1,117.53	-440.95	-308.60
NAVEL	3.48	1.78	0.61	-2.96	-0.68	0.51
COLOR	16.25**	5.09**	-1.86	2.37	0.73	-0.13
FLY	-3.44	0.62	1.16	-2.38	-0.56	0.02
TICK	0.49	-0.80	0.30	-10.34	-0.68	2.53
COAT	2.78	0.75	-1.43	0.16	2.34	-1.32
COND	-4.61*	-2.26**	-0.60	2.23	-0.60	0.74
YWT	189.02**	38.80	4.58	-15.10	-108.42**	30.61

*P < 0.05; **P < 0.01

CONCLUSIONS

Traditionally, inbreeding coefficients – based on the probability that both alleles at any given locus within an individual are identical by descent – have been computed from pedigree information. However, the advent of high-density genotype data allows for the estimation of realised F which might differ even between animals that have identical pedigree.

It is worth mentioning that the pedigree-based inbreeding was of no practical use in our case because after tracing back three generations of ancestors, only four Brahman individuals were found to be inbred (all with F = 12.5%) and no Tropical Composite individuals were inbred. This is not surprising as the Beef CRC cattle were experimental cattle representing many properties, and bred with the objective to avoid, or at least minimise inbreeding.

While the average of the three inbreeding measures employed here were similar in the two populations, the Brahman population appeared to be more affected by inbreeding depression than the Tropical Composite. A comparison of the inbreeding depression estimates here with those obtained from a model that contains a single inbreeding measure (as opposed to the three of them), plus a random additive polygenic component is warranted.

Also in the future, we plan to explore different parameterizations of the GRM, and to partition ROH into two groups – 'long' and 'short,' with a view to infer past *versus* recent inbreeding as described by Saura *et al.* (2015). Characterising the gene content in the regions prioritised by differential ROH could also help identify genes of commercial relevance in the two populations. Finally, it would be worthwhile to establish what relationship exists between inbreeding and the 'missing heritability' problem, as this has still not yielded to analysis.

REFERENCES

Keller M.C., Visscher P.M. and Goddard M.E. (2011) *Genetics* **189**: 237.

Porto-Neto L., Reverter A., Prayaga K.C., Chan E.K.F., Johnston D.J. *et al.* (2014) *PLoS ONE* **9**: e113284.

Pryce J.E., Haile-Marian M., Goddard M.E. and Hayes B.J. (2014) *Genet. Sel. Evol.* **46**:71.

Saura M., Fernandez A., Varona L., de Cara M.A.R., Barragan C. *et al.* (2015) *Genet. Sel. Evol.* **47**: 1.

Van Raden PM (2008) *J. Dairy Sci*. **91**: 4414.

Proc. Assoc. Advmt. Breed. Genet. 21: 449-452

OPTBR: COMPUTATIONALLY EFFICIENT GENOMIC PREDICTIONS AND QTL MAPPING IN MULTI-BREED POPULATIONS

Tingting Wang[1,2,3], Yi-Ping Phoebe Chen[1], Kathryn E. Kemper[4], Michael E. Goddard[2,3,4] and Ben J. Hayes[1,2,3]

[1] Faculty of Science, Technology and Engineering, La Trobe University, Victoria, Australia
[2] AgriBio, Centre for AgriBioscience, Biosciences Research, DEDJTR, VIC 3083, Australia
[3] Dairy Futures Cooperative Research Centre, VIC 3083, Australia
[4] Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Victoria, Australia

SUMMARY

As genomic data used for prediction of complex traits rapidly expand in size, the importance of computational efficiency of genomic prediction algorithms becomes paramount. In this paper we describe an expectation-maximisation (EM) algorithm for genomic prediction (OptBR) with the speed-up scheme that is up to 30 times faster than MCMC implementations. The algorithm is flexible for joint analysis of data from different sources, as it includes weightings for the accuracy of phenotype, and can accommodate effects of factors such as breed, age, sex and additional covariates. A further advantage of the method is that QTL mapping is performed simultaneously with genomic prediction.

INTRODUCTION

Genomic predictions are increasingly used to identify breeding individuals in livestock and crop improvement programs. The prediction equation to calculate genomic predictions is derived from a reference population genotyped for thousands of single nucleotide polymorphisms (SNPs), and with phenotypes for the target trait (Meuwissen *et al.* 2001), or through an alternative implementation where genomic relationships derived from the SNP are used to predict breeding values for selection candidates (e.g. VanRaden 2008). Across many species, a key finding is that reference populations must be very large to achieve high accuracies of genomic prediction. One way to increase the size of the reference population is to combine information across populations from the same species. For example in dairy and beef cattle small to moderate increases in prediction accuracy have been reported by using a multi-breed reference population (Lund *et al.* 2014; Kemper *et al.* 2015; Bolormaa *et al.* 2013). Another finding from these studies is that the increase in accuracy of prediction from combining information across populations can depend on the method of prediction.

For multi-breed predictions, methods which assume *a priori* that SNP effects are all non-zero and small, and all from the same normal distribution (SNP-BLUP and GBLUP) do not perform as well as methods that assume *a priori* that some SNP may have zero, small or moderate to large effects (BayesB, or BayesR) (Lund *et al.* 2014; Kemper *et al.* 2015). Compared to BLUP methods, these models use priors which assume a large proportion of SNP have effects close to zero, or actually zero, while a small proportion of SNP have moderate to large effects. This is important not only to improve genomic predictions across breeds, but also to improve the precision of QTL mapping using such methods. While the Bayesian methods are very attractive, the major difficulty with these methods is long computation time, which becomes intractable with very large data sets. The long computational time arises because Bayesian methods are typically implemented using MCMC. To speed up Bayesian methods, several heuristic convergence methods have been proposed e.g. fastBayesB (Meuwissen *et al.* 2009) or fastBayesA (Sun *et al.* 2012). All of these methods reported reduced computation time but in some cases the prediction accuracy was reduced compared to their MCMC counterparts.

Our aim was to develop a computationally efficient algorithm (OptBR for Optimized BayesR) for simultaneous multi-breed prediction and QTL mapping. OptBR implements an EM algorithm on the hierarchical prior assumption for SNP effects and other parameters from BayesR (Erbe *et al.* 2012). Also, OptBR retains the advantage of Predicted Error Variance (PEV) correction of emBayesR (Wang *et al.* 2015) to improve the accuracy. OptBR has four improvements compared with emBayesR which allow it to be applied to very large data sets, which may encompass multiple populations. These advantages include 1) weighting of phenotypes to allow for different errors in measurement across populations; 2) multi-breeds are accounted for by introducing fixed effects into the prediction models; 3) a polygenic term to capture variation not explained by the SNP, and 4) a speed-up scheme to make it 30 times faster than BayesR implemented with MCMC.

MATERIALS AND METHODS

Genotypes and phenotypes. OptBR was implemented on 630K SNPs panels (with total 632,003 SNPs), that was imputed from 777K and 54K Illumina Bovine SNP arrays. Phenotypes for milk yield, protein yield, fat% and fertility were daughter trait deviations (DTD) for bulls, and trait deviations (TD) for cows. For genomic prediction, the data was separated into references set and validation sets. The reference data included 16,214 Holstein and Jersey bulls and cows, while the validation set included 251 Holstein bulls (bulls born after 2007), or a third breed, 114 Australian Red bulls (Australian reds bulls were never included in the reference set).

Data Model. The statistical model is $\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}\text{g} + \mathbf{Wv} + \mathbf{e}$ where $\boldsymbol{\beta}$ is a vector of fixed effects including breed, $\mathbf{g}$ is a vector of the SNP effects, $\mathbf{v}$ is a vector of polygenic effects ~ $N(0, \mathbf{A}\sigma_v^2)$, $\mathbf{e}$ is a vector of residuals ~ $N(0, \mathbf{E}\sigma_e^2)$ where $\mathbf{E}$ is diagonal and accounts for error in TD and DTD, with σ_e^2 the error variance. Three design matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{W}$ allocate phenotype ($\mathbf{y}$) to the vectors $\boldsymbol{\beta}$, $\mathbf{g}$, and $\mathbf{v}$ separately. The SNP effects are assumed to be drawn from a mixture of normal distributions with zero mean and variance either $0 \ or\ 0.0001 * \sigma_g^2 \ or\ 0.001 * \sigma_g^2 \ or\ 0.01 * \sigma_g^2$ with probability $\mathbf{Pr}_{\boldsymbol{k}}$ ($k = 1 \cdots 4$) drawn from a Dirichlet distribution with parameters (1,1,1,1).

Expectation maximisation algorithm. To implement the EM algorithm we rewrite the statistical model for the ith SNP as $\mathbf{y} = \mathbf{X\beta} + \mathbf{Z_i}\text{g}_\text{i} + \mathbf{u_1} + \mathbf{Wv} + \mathbf{e}$ where $\mathbf{u_1} = \boldsymbol{Zg} - \mathbf{Z_i}\text{g}_\text{i}$, that is $\mathbf{u_1}$ is the sum of all SNP effects other than SNP *i*. This form of the model allows us to treat $\mathbf{u_1}$ as missing data and take expectations of the likelihood over $\mathbf{u_1}$ and hence estimate g_i allowing for the errors in the estimates of all the other SNP effects. We take the expectation of the log Likelihood of $\mathbf{y}$ using $Var(\mathbf{u_1}|\mathbf{y}) = \mathbf{PEV}(\mathbf{u_1})$ where the prediction error variance (PEV) is derived from a BLUP approximation to the mixture model. We then maximize the expected likelihood with respect to each of the parameters including g_i, the mixing proportions ($\mathbf{Pr}$), $\boldsymbol{\beta}$ and $\mathbf{v}$ as well as σ_e^2. We also trialled a speed-up scheme: when the SNP effect g_i is very small ($|\text{g}_\text{i}| \geq 0.00000001$) after 50 iterations, it was not updated in future iterations but left at its current value.

RESULTS AND DISCUSSION

To compare computing times for OPTBR and BayesR, three reference data sets related to milk yield were used, which have 632,003 SNPs with different numbers of animals ranging from 3,049 in RefI (Holstein bulls Only), 11,527 in RefII (Holstein bulls and cows), to 16,214 in RefIII (Holstein and Jersey bulls and cows) seen in Figure 1. The results demonstrate the advantage of OptBR over BayesR, and the advantage of the speed-up scheme. For instance, in the largest dataset time to convergence was 720 hours for BayesR but 28 hours for OptBR_Sp.

The accuracies of prediction using the EM were similar to BayesR with the exception of fat% (Table 1). A detailed investigation of the speed-up scheme was assessed using milk yield. Table 1 shows that the speed up procedure did not sacrifice any accuracy (Table 1).

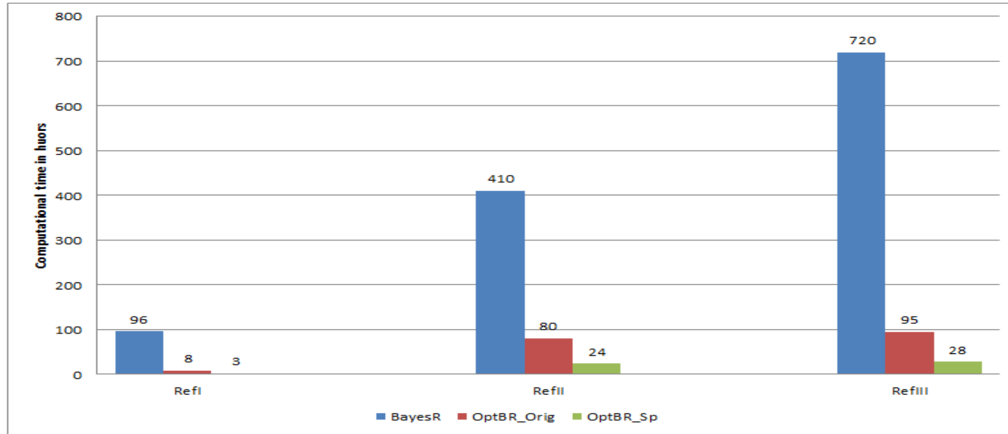


Figure 1. The computational time in hours compared between BayesR, OptBR_Orig, and OptBR_Sp on three reference data sets (RefI with 3,049 animals, RefII with 11,527 animals, and RefIII with 16,214 animals)**.**

Table 1. The impact of the speed-up scheme C1 on accuracy (Acc.), the proportion of variants in each distribution (Pr) and error variance (σ_e^2) using milk yield as an example.

	Acc.	Pr	σ_e^2
OptBR_Orig	0.66	[0.998371, 0.001583, 0.000007, 0.000039]	239409
OptBR_Sp	0.68	[0.997545, 0.002394, 0.000009, 0.000052]	247965

The results in Table 2 demonstrate the robust prediction ability of our algorithm OptBR for multi-breeds and across breed prediction. On milk production traits, both BayesR and OptBR have 3%~7% advantage over GBLUP. On the fertility, three methods had the similar performance. The prediction accuracy for Australian red bulls was not as high as for Holstein, which is not surprising given there were no Australian Reds in the data set. The bias is the coefficient of regressing the phenotype of validation set on Genomic Estimated Breeding Value (GEBV), which shows the underestimation of three methods for SNP effects on most of the traits except Fertility.

Table 2. The accuracy (Acc.) and bias of predictions for BayesR, GBLUP and OptBR from the Holstein and Jersey multi-breed reference population using either the Holstein or Australian Red validation populations.

	Milk Yield		Protein Yield		Fat%		Fertility	
	Acc.	Bias	Acc.	Bias	Acc.	Bias	Acc.	Bias
Holstein validation								
BayesR	0.68	0.84	0.68	0.88	0.81	0.90	0.44	1.53
GBLUP	0.63	0.83	0.65	0.85	0.74	0.85	0.44	1.66
OptBR	0.68	0.90	0.68	0.79	0.77	0.83	0.44	1.27
Australian Reds validation								
BayesR	0.22	0.60	0.12	0.49	0.45	0.92	0.27	1.03
GBLUP	0.16	0.54	0.11	0.51	0.32	0.90	0.29	0.97
OptBR	0.24	0.70	0.12	0.42	0.41	0.89	0.29	1.10

We compared the ability of BayesR and OptBR to map QTL by investigating the number of SNPs with high posterior probabilities of having a non-zero effect (Figure 2). The number and position of QTL was similar between BayesR and OptBR. For milk yield, similar to BayesR, OptBR finds SNPs near to the genes *CSF2RB* located on chromosome 5, SNPs near the casein complex on chromosome 6 (~87Mb), and SNPs related to *CCL28/GHR* on chromosome 20. The well-known gene *DGAT1* (on chromosome 14) is mapped by both BayesR and OptBR.

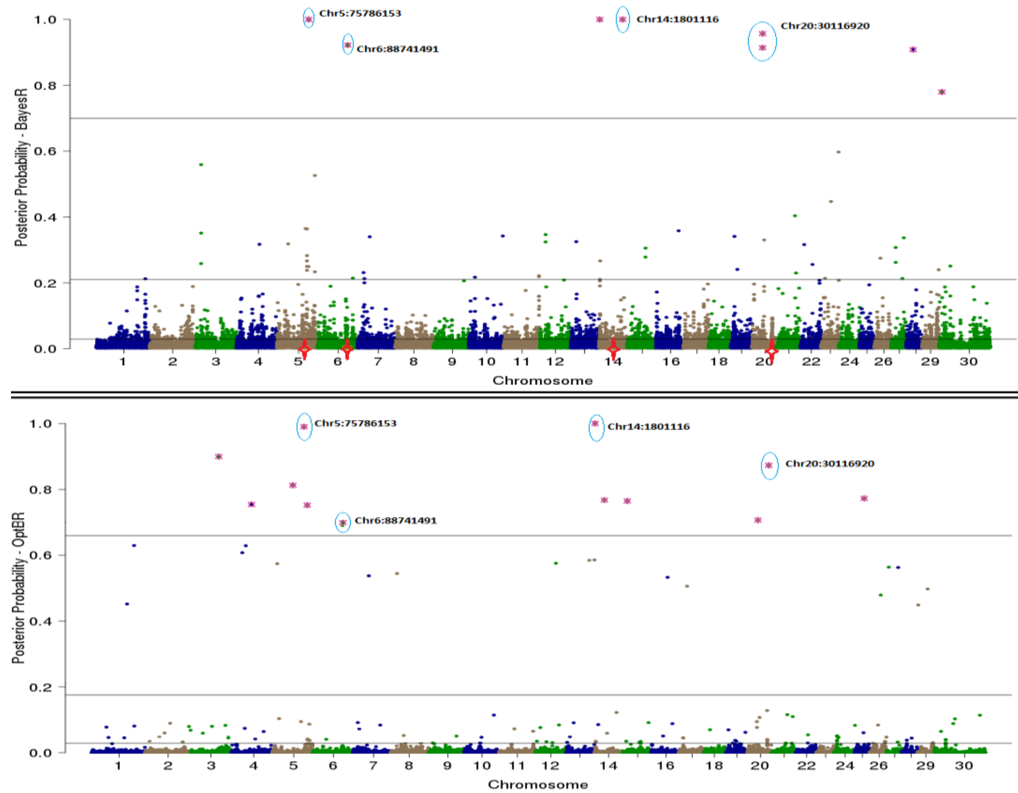


Figure 2. Posterior probability of non-zero SNP effect for milk yield from BayesR (top) and OptBR (bottom) across all chromosomes.

The results suggest that OptBR will be useful for simultaneous genomic prediction and QTL mapping, particularly for very large data sets where computational efficiency is very important.

REFERENCES

Meuwissen T.H.E., Hayes B.J. and Goddard M.E. (2001) *Genetics.* **157**:1819.
VanRaden P.M. (2008) *J Dairy Sci*. 91(11):4414-23.
Lund M.S., Su G., Janss L., Guldbrandtsen B. *et al.* (2014) *Livest Sci.* **166:**101.
Yang J., Benyamin B., McEvoy B.P, Gordon S., *et al.* (2010). *Nat Genet.* **42:**565.
Kemper K.E., Reich C.M., Bowman P.J., Vander Jagt C.J., *et al.* (2015) *Genet Select Evol*. **47**:29.
Bolormaa S., Pryce J.E., Kemper K., Savin K., *et al.* (2013) *J Anim Sci*. **91**:3088.
Meuwissen T.H., Solberg T.R., Shepherd R. and Woolliams J.A. (2009) *Gen Sel Evol*. **41**:2.
Sun X., Qu L., Garrick D.J., Dekkers J.C.M. *et al.* (2012) *PLOS ONE.* **7(11)**: e49157.
Wang T., Chen Y.P.P., Goddard M.E., Meuwissen T.H. *et al. Gen Sel Evol*. **47**:34
Erbe M., Hayes B.J., Matukumalli L.K., Goswami S., *et al.* (2012) *J. Dairy Sci*. **95**:4114.

Proc. Assoc. Advmt. Breed. Genet. 21: 453-456

THE IMPACT OF FITTING INCORRECT MODELS ON THE PARTITIONING OF GENETIC VARIANCE COMPONENTS FOR BIONOMIAL LAMB TRAITS

P.R. Amer K. Stachowicz, P. Pierard and G.M. Jenkins

AbacusBio Limited, Dunedin, New Zealand

SUMMARY

This paper presents a simulation study that shows that failure to fit dam permanent environmental effects in variance component estimation of lamb survival results in an upward bias in the estimate of maternal genetic variance. In contrast, fitting litter effects has little impact on variance component estimation. These results have implications for sheep genetic evaluation of lamb survival, and values for the direct and maternal heritability of lamb survival at the lower end of the range of those published based on alternative models should be used in national genetic evaluations.

INTRODUCTION

Lamb survival is lowly heritable, with many sources of variation affecting phenotype. These sources of variation include environmental effects, dam lifetime permanent environmental effects, litter effect, maternal genetic effects and direct genetic effects. Concerns have been raised that data structure may be insufficient for genetic variance component estimation software to partition maternal genetic effects from dam permanent environmental effects, meaning that some of the variance explained by maternal genetics may be incorrectly assigned to dam permanent environmental effects. This could lead to the, perhaps false, assumption that the maternal genetic effect is not significant. The opposite could also be true, and where permanent environmental effects are not accounted for, the maternal genetic effect could be artificially inflated. The aim of this study was to determine whether the data structures typical in New Zealand sheep flocks are sufficient for partitioning maternal genetic variance from environmental litter variance and dam lifetime permanent environmental variance. The distribution of the trait was also tested to determine if there was an effect on variance component estimation due to the trait being binary instead of continuous, and when binary, for dependence on incidence.

MATERIALS AND METHODS

This study was based on simulation of a population of animals born during 1995-2001. The population structure and relationship between individuals is set based on a real pedigree from a single sheep flock to reflect industry data, so as to be typical of that which would be used for variance component estimation in practice. The proportions of ewes lambing singles, twins, triplets and greater litter sizes were .15, .68, .16 and .01 respectively. There were 103 sires with at least progeny 30 progeny, and a total of 20,107 lambs born across the seven years. There was excellent connectedness across years due to several rams having large numbers of progeny across many years, and a further 18 rams with progeny across either 2 or 3 years (see Amer and Jopson, 2003, for further details on the pedigree structure).. Animals from the first generation (i.e. unknown parentage) were attributed random true breeding values, as sampled from a normal distribution with a mean of 0 and genetic variances (direct and maternal) equal to the respective heritabilities (total phenotypic variance was assumed to be 1). As lamb survival is a binary trait but phenotypes were sampled from a normal distribution, a threshold (described below) was set to determine survival on the binary scale. The lamb survival phenotypes of individual progeny of known parents identified in the existing pedigree in subsequent generations were derived using methods described below.

Five random effects were simulated as contributing to the continuous lamb survival phenotype (aphen): the direct genetic effect, maternal genetic effect, dam permanent environmental effect, litter environmental effect and a residual effect. All founders in the flock pedigree were assigned values for genetic parameters, and then the simulation iterated down through generations within the pedigree simulating genotypes and phenotypes for all animals present, based on values previously simulated for their parents. To preserve pedigree integrity, no removal of animals as parents was undertaken, even when their simulated phenotype was "dead".

Two different binary (alive or dead) traits were derived from the continuous phenotype – one to represent high and one to represent low survival rates. To simulate a high survival binary trait, a threshold of 1.4 was defined to set survival to either dead (>1.4) or alive (≤1.4) – equating to approximately 8% of lambs in the population dying. A threshold of 1 was used to simulate a low survival binary trait - corresponding to approximately 15% lamb death. Expected values for heritabilities of binary traits when estimated using normal linear models was derived based on Dempster and Lerner (1950).

Six scenarios with varying levels of simulated variance components were used to simulate each of the lamb survival traits (Table 1). HIGH (Table 1) reflects a situation where there are larger variance components in the lamb survival trait. LOW (Table 1) is perhaps more reflective of the real world scenario, where direct and maternal heritabilities are low and dam permanent environmental effect is low and equivalent to direct and maternal heritabilities. LOW1-4 (Table 1) were variations of LOW in that the same amount of genetic variation was present within the trait with each scenario (exception - LOW3), but the variance was distributed among the parameters differently - LOW1 (no litter effect), LOW2 (no permanent environmental effect), LOW3 (no litter or permanent environmental effect) and LOW4 (no maternal genetic effect). Each scenario was simulated 20 times and results averaged.

For each trait (continuous, low survival, high survival) and each simulation scenario (HIGH, LOW, LOW1, LOW2, LOW3 and LOW4) five different models were fitted in ASREML (Gilmour et al, 2008) (Table 2). Although lamb survival is a binary trait, the continuous version was also analysed to determine if the efficiency of estimation and partition of variance components was influenced by the distributional properties of the trait.

Table 1 Alternative sets of variance component parameters used as simulation inputs

	Scenario					
	HIGH	**LOW**	**LOW1**	**LOW2**	**LOW3**	**LOW4**
h^2direct	0.2	0.05	0.05	0.05	0.05	0.05
h^2maternal	0.15	0.05	0.05	0.05	0.1	0
damPE (c^2)	0.1	0.05	0.1	0	0	0.1
litter $(l)^2$	0.05	0.05	0	0.1	0	0.05

RESULTS AND DISCUSSION

When dam permanent environmental effect (damPE) is present in the dataset and accounted for in the genetic analysis, the estimate of maternal heritability of lamb survival is what would be expected based on the known (simulated) values in the dataset (Table 3). That is, the data structure is sufficient for the partitioning of variance between maternal genetics and dam permanent environmental effects. Conversely, when dam permanent environmental effects are present in the data but are not accounted for in the statistical model fitted, then the maternal heritability is artificially inflated (Table 3). This is the case whether a continuous lamb survival phenotype

(0.091 when damPE is not fitted versus 0.049 when damPE is fitted), low survival phenotype (0.039 versus 0.021) or high survival phenotype (0.029 versus 0.016) is simulated. These results were consistent with results from the model when applied to the HIGH variance component set, whereby all components (direct genetics, maternal genetics, dam permanent environmental effects and litter effects) are simulated to be present within the population (Table 4). In this scenario, failure to account for dam permanent environmental effect resulted in a 35-55% increase in maternal heritability (Table 4).

Table 2 Mixed models fitted - models 1-4 are fitted for the binary and continuous traits, while the probit model (5) is only fitted for the binary traits

Model	Equations
1	cSurv/lSurv/hSurv = byr + sex + br + aod + animal + dam + dampe + litter
2	cSurv/lSurv/hSurv = byr + sex + br + aod + animal + dam + dampe
3	cSurv/lSurv/hSurv = byr + sex + br + aod + animal + dam + litter
4	cSurv/lSurv/hSurv = byr + sex + br + aod + animal + dam
5 (probit)	lSurv/hSurv (probit) = byr + sex + br + aod + animal + dam + dampe + litter

cSurv = continuous lamb phenotype, lSurv = low incidence lamb survival binary phenotype, hSurv = high incidence lamb survival binary phenotype, byr = birth year, br = birth rank, animal = animal's unique identifier (relationships included by fitting the A matrix, which describes pedigree relationships, dam = dam uid, dampe = permanent environmental effect, litter = litter effect.

Table 3 Effect of not accounting for permanent environmental effect (damPE) when it is present in the data (low survival variance assumed) – expected based on simulated values and estimated based on variance component estimation

TRAIT	h^2direct		h^2maternal		damPE	
	Expected	estimated	expected	estimated	expected	estimated
Dam permanent environmental effect simulated and fitted						
Continuous	0.05	0.055	0.05	0.049	0.1	0.042
Low survival	0.021	0.025	0.021	0.021	0.1	0.021
High survival	0.015	0.015	0.015	0.016	0.1	0.017
Dam permanent environmental effect simulated but not fitted						
Continuous	0.05	0.054	0.05	0.091	0.1	n/a
Low survival	0.021	0.025	0.021	0.039	0.1	n/a
High survival	0.015	0.016	0.015	0.029	0.1	n/a

Table 4 Effect of not accounting for either permanent environmental effect (damPE) or litter effect when they are present in the data (high survival variance assumed) – expected (exp) based on simulated values and estimated (est) based on variance component estimation

TRAIT	h^2direct		h^2maternal		damPE		litter	
	exp	est	exp	est	exp	est	exp	est
All terms simulated, all terms fitted								
Continuous	0.2	0.203	0.15	0.153	0.1	0.047	0.05	0.010
Low survival	0.086	0.097	0.064	0.065	0.1	0.033	0.05	0.009
High survival	0.061	0.072	0.046	0.050	0.1	0.030	0.05	0.013
All terms simulated, litter not fitted								
Continuous	0.2	0.203	0.15	0.153	0.1	0.050	0.05	n/a
Low survival	0.086	0.097	0.064	0.065	0.1	0.035	0.05	n/a
High survival	0.061	0.072	0.046	0.050	0.1	0.033	0.05	n/a
All terms simulated, dam permanent environmental effect not fitted								
Continuous	0.2	0.193	0.15	0.206	0.1	n/a	0.05	0.013
Low survival	0.086	0.094	0.064	0.099	0.1	n/a	0.05	0.013
High survival	0.061	0.071	0.046	0.078	0.1	n/a	0.05	0.016

Additional simulations (results not shown) revealed that when variance components were not simulated, they were correctly predicted as being zero. Furthermore, when a generalised linear model approach was used to account for binomially distributed data, the only difference in results was as could be predicted using the transformation described by Dempster and Lerner (1950).

Everett Hincks et al (2014) published variance component estimates for NZ sheep using a range of variance models. One of their models included dam permanent environmental effects, and this model gave the lowest estimate of the maternal genetic variance of lamb survival. We conclude that typical sheep pedigree structures in New Zealand sheep populations are sufficient for disentanglement of multiple variance components, and failure to fit an existing random effect in variance component estimation procedures can result in inflation of other variance components.

REFERENCES

Amer, P.R. and Jopson, N.B. (2003). *Proc. NZ Soc. Anim. Prod.*, **63:** 201.

Dempster, E.R. and Lerner, I.M. (1950). *Genetics*, **35:** 212.

Everett Hincks, J.M., Mathias-Davis, H.C., Greer, G.J., Auvray, B.A. and Dodds, K.G. (2014). *J. Anim. Sci.*, **92(7):** 2885.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. (2009). ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Proc. Assoc. Advmt. Breed. Genet. 21: 457-460

BACKFAT AS AN ENVIRONMENTAL DESCRIPTOR IN DEFINING GROWTH RATE OF THE PIG: A G×E ANALYSIS

S.Z.Y. Guy[1], S. Hermesch[2] and P.C. Thomson[1]

[1]ReproGen Animal BioScience Group, Faculty of Veterinary Science, University of Sydney, NSW, Australia
[2]Animal Genetics and Breeding Unit, a joint venture of NSW Department of Primary Industries and University of New England, University of New England, Armidale, NSW, Australia

SUMMARY

Investigations of genotype by environment (G×E) interactions may use estimates of average performance observed for contemporary groups (CGs) as environmental descriptors (ED). Data from a commercial breeding herd of Large White pigs were used to define ED based on backfat (BF) and average daily gain (ADG). The ED of BF and ADG were estimated using an animal model, with sex, month-year CG, weight (BF only), litter size and parity of birth litter as fixed effects. Estimates of CG were centred, and then used to allocate an environment for each individual in the genetic analyses of ADG. Each ED was partitioned into quartiles, allowing ADG to be defined as a separate trait in the four environments based on BF or ADG. Heritability estimates for ADG ranged from 0.12 to 0.16 for BF as ED, and 0.07 to 0.17 for ADG as ED. There was a weak relationship between the BF ED and ADG ED indicating they do not quantify the environment in the same way. Nevertheless, the use of either ED indicates re-ranking of animals in different environments, with Pearson's correlations between EBVs ranging from 0.22 to 0.55 for BF as ED, and 0.43 to 0.54 for ADG as ED.

INTRODUCTION

Genotype by environment (G×E) interactions occur when different genotypes exhibit varying responses to changes in the environment. Phenotype, in particular mean performance of a group of animals, can be seen as the result of a combination of known, plus unknown and unobservable environmental factors (Streit *et al.* 2013). Therefore, estimates of phenotypic averages of contemporary groups (CGs) at each environmental level are commonly used as an environmental descriptor (ED) in animal breeding, allowing the environment to be quantified (for example, Knap and Su (2008) in pigs). This ED can then be partitioned, and the same trait measured in the different EDs can then be considered as separate traits (Falconer 1952), with each trait having its own heritability and breeding values. This multi-trait approach of G×E analysis allows the evaluation of any genetic correlations (r_g) between the same trait expressed in different environments, and, if less than unity, this indicates a G×E interaction.

In pig breeding, an environmental variable previously used was average daily gain (ADG) (Li and Hermesch 2013).We explore the use of backfat (BF) as an alternate production trait for an ED in G×E analyses, and make comparisons with the use of ADG as the ED.

MATERIALS AND METHODS

Data. Pig identity records and production traits were obtained from a commercial herd of Large White pigs in Gatton, south east Queensland, Australia. Inclusion criteria were years of birth from 1996 to 2013 inclusive, and all traits within four standard deviations from means of the raw data. After data editing, there were a total of 40,145 individual animals, which included 19,899 entire male pigs and 20,246 female pigs. The 18 generations consisted of 2,444 dams and 568 sires. Performance traits included ADG from birth to weighing and BF at weighing. The mean age of weighing was 129.1 ± 6.79 days (mean ± SD), which gave an average weight at testing of 87.1

± 9.23 kg. CGs were defined by birth month-year, giving a total of 216 CGs, ranging from 67 to 493 pigs in each group and an average group size of 185 pigs.

Analysis. Data cleaning and analysis was conducted using R, version 3.1.3 (R Core Team 2015). Models were fitted using ASReml-R (Butler *et al.* 2009). Records were examined for duplicates and errors. The pedigree was extracted from the raw records, duplicate pigs and pedigree loops were removed, and founders identified. The analyses were conducted in two steps.

Firstly, EDs were obtained based on estimates of CGs from the following animal models. The model for BF was BF = μ + Sex + CG + βWeight + βLitterSize + BirthParity + Animal + Litter effect +ε. Fixed effects were sex, CG, weight (linear covariate), litter size of birth litter (linear covariate) and parity of birth litter. Random effects were common litter and animal effect. For ADG as the ED, the model was: ADG = μ + Sex + CG + βLitterSize + BirthParity + Animal + Litter effect + ε.

The 216 CG estimates for both EDs were centred around 0, and for maximum power to test for G×E, split into quartiles to have roughly equal number of observations within each group. Each animal was allocated an environment (E-BF1, E-BF2, E-BF3, or E-BF4; as well as E-ADG1, E-ADG2, E-ADG3, or E-ADG4) according to their CG estimate.

The second part of the analyses was to define ADG as a different trait for each environmental group. Heritabilities and estimated breeding values (EBVs) for ADG traits across environments were obtained from the animal model outlined for ADG above. Pearson's correlations between the EBVs for each of the four traits based on BF as ED, as well as ADG as ED, were calculated as a proxy measure of genetic correlations.

RESULTS AND DISCUSSION

The 40,145 animals included in analysis had a mean ADG of 675.3 ± 68.43 g/day, and a mean BF measurement of 11.6 ± 1.90 mm.

The centred CG estimates derived from the animal models in the first step of analysis ranged from -1.2 mm to 1.3 mm for the BF ED, and from -67.2 g/day to 77.5 g/day for the ADG ED. The environments E-BF1, E-BF2, E-BF3 and E-BF4 contained animals with a BF ED of < -0.38 mm, between -0.38 mm and 0.01 mm, between 0.01 mm and 0.39 mm, and > 0.39 mm, respectively; Similarly, E-ADG1, E-ADG2, E-ADG3 and E-ADG4 contained animals with an ADG ED of < -15.9 g/day, between -15.9 g/day and 1.34 g/day, between 1.34 g/day and 16 g/day, and > 16 g/day, respectively.

In an optimum environment, pigs have a higher ADG and lower BF. If the BF ED and ADG ED were able to quantify the environment in the same way, it was expected for these EDs to be highly negatively correlated. Figure 1 shows the weak relationship between the EDs based on BF and ADG (r = 0.08). This indicates that the two EDs do not describe the environment in the same way.

The partitioning of the environments appropriately described inferior and superior environments, shown in the ADG of each environment. The superior BF environments with the lowest BFs had the highest ADG performance, with ADG decreasing from 680 g/day and 681 g/day for E-BF1 and E-BF2, to 668 g/day and 671 g/day for E-BF3 and E-BF4 (Table 1). The ED derived from ADG showed an increase in ADG with quality of environment in a linear relationship, as expected. Variability in performance (CV) decreased with superior environments for ADG as ED, reflecting the results of Li and Hermesch (2013) for their seven-trait analysis. The range of heritabilities derived from the four ADG traits in each ED were 0.12 to 0.16 for ED based on BF, and 0.07 to 0.17 for ED based on ADG.

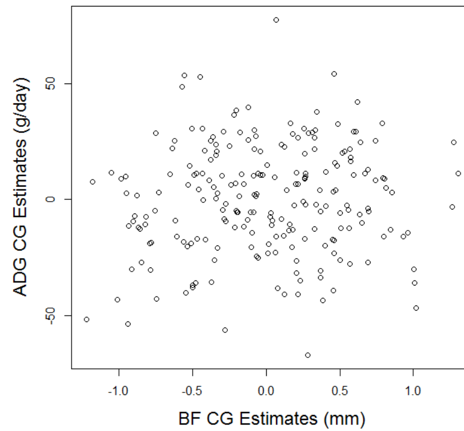


Figure 1. The relationship between centered contemporary group (CG) estimates using backfat (BF) as ED and average daily gain (ADG) as ED (r = 0.08).

Table 1. Number of observations (n), mean performance, coefficient of variation (CV), phenotypic variance (σ_p^2), heritability (h^2), standard error of heritability estimate (s.e(h^2)), fraction of variance due to common litter environment (c^2) and litter effect standard error ((s.e(c^2)) for average daily gain (ADG) defined as separate traits, using an environmental descriptor (ED) derived from backfat (E-BF1 to E-BF4) and ADG (E-ADG1 to E-ADG4).

Environment	n	ADG (g/day)	CV (%)	σ_p^2	h^2	s.e(h^2)	c^2	s.e(c^2)
E-BF1	9,767	680.0	9.97	3948.2	0.16	0.027	0.10	0.025
E-BF2	11,328	680.6	10.07	4143.7	0.16	0.025	0.09	0.026
E-BF3	9,804	668.4	10.33	4068.8	0.12	0.022	0.12	0.028
E-BF4	9,246	671.2	10.04	4110.7	0.14	0.025	0.12	0.026
E-ADG1	9,924	648.3	10.06	3941.6	0.15	0.027	0.11	0.025
E-ADG2	10,313	670.8	9.71	4034.8	0.17	0.026	0.11	0.024
E-ADG3	10,695	682.8	9.68	4158.7	0.07	0.018	0.11	0.019
E-ADG4	9,213	700.7	9.56	4133.6	0.17	0.026	0.13	0.026

Pearson's correlations between EBVs ranged from 0.22 to 0.55 for BF as ED, and from 0.43 to 0.54 for ADG as ED (Table 2). These were all significantly lower than unity, demonstrating re-ranking of animals across environments. Although Pearson's correlations indicate significant G×E interactions for both BF as ED and ADG as ED, these provisional values under-estimate genetic correlation between traits.

Table 2. Pearson's correlations between estimated breeding values (EBVs) for average daily gain (ADG) defined as separate traits in each environment, using an environmental descriptor (ED) derived from (a) backfat (BF) and (b) ADG.

(a)

	E-BF1	E-BF2	E-BF3	E-BF4
E-BF1				
E-BF2	0.35			
E-BF3	0.29	0.55		
E-BF4	0.22	0.49	0.54	

(b)

	E-ADG1	E-ADG2	E-ADG3	E-ADG4
E-ADG1				
E-ADG2	0.53			
E-ADG3	0.43	0.52		
E-ADG4	0.45	0.50	0.44	

This multi-trait approach treats the ED as a categorical variable. When the ED is treated as a continuous variable, a reaction norm (RN) approach can be used (Kolmodin 2003). There is also the option of combining both approaches, when both categorical and continuous EDs are used at the same time. Windig *et al.* (2011) explored treatment of the ED as both continuous and categorical in a combined bivariate reaction norm approach. Although there was no G×E interaction found when multi-trait, RN and combined approaches were used, the combined approach was useful for separating effects when two EDs were confounded (e.g. spring calving vs. year-round calving production system). In this example, residual variance decreased with dairy higher milk production in a RN approach, but the combined approach showed that at the same milk production level, there was higher residual variance in spring calving compared to year round calving.

The number of traits the environmental trajectory is split into is an important factor in G×E analysis. Li and Hermesch (2013) explored four different scenarios, splitting ADG as ED into one, two, three and seven traits. When treated as one and two traits, no significant G×E interaction was found, but a G×E interaction was observed when three and seven trait models were fitted. Genetic correlations also decreased as differences between environmental groups increased. Quartiles were used in the current study as it is a commonly used statistical summary. However the optimum number of traits should be further investigated.

CONCLUSIONS

This paper considers the validity and feasibility of G×E analyses when using alternative traits in defining the environmental variable. The mean performance of a production trait as ED, adjusted for by fixed and random effects, may be an appropriate variable if the environment is complex, or if there is no other available data to describe the environment. These first results indicate that BF can be used as an ED, with estimates of heritabilities and Pearson's coefficients similar to those obtained when ADG was used as the ED. Both EDs suggest re-ranking of animals across environments. However, genetic correlations between ADG defined as a separate trait in different environments are required to make a final conclusion about G×E interactions.

ACKNOWLEDGEMENTS

S.Z.Y. Guy is the recipient of a Pork CRC postgraduate scholarship.

REFERENCES

Butler D.G., Cullis B.R., Gilmour A.R. and Gogel B.J. (2009) Queensland Department of Primary Industries and Fisheries, Queensland.

Falconer D.S. (1952) *Amer. Nat.* **86**: 293.

Knap P.W. and Su S. (2008) *Animal* **2**:1742.

Kolmodin R. (2003) Doctoral Thesis, Swedish University of Agricultural Sciences.

Li L. and Hermesch S. (2013) *Proc. Assoc. Advmt. Anim. Genet.* **20**:323.

R Core Team (2015). *R Foundation for Statistical Computing,* Vienna, Austria.

Streit M., Wellmann R., Reinhardt F., Thaller G., Peipho H-P. *et al.* (2013) *G3*. **3**:1085.

Windig J.J, Mulder H.A., Bohthe-Wilhelmus D.I. and Veerkamp R.F. (2011) *J. Dairy Sci.* **94**:3137.

Proc. Assoc. Advmt. Breed. Genet. 21: 461-464

WHICH GENOMIC RELATIONSHIP MATRIX?

B. Tier[1], K. Meyer[1] and M. H. Ferdosi[1,2]

[1]Animal Genetics and Breeding Unit*, University of New England, Armidale NSW 2351.
[2]School of Environmental and Rural Science, University of New England, Armidale NSW 2351.

SUMMARY

Genomic information can accurately specify relationships among animals, including between those without known common ancestors. Genetic variances estimated with genomic data relate to unknown, more distant, founder populations than those defined by the pedigree. Starting from different sets of assumptions, the properties of some alternative genomic relationship matrices (**G**) are explored. Although the assumptions and matrices differ, the resulting sets of estimated breeding values predict the differences between animals identically, despite obtaining different estimates of the additive genetic variance – showing that there are many ways of building **G** that provide identical results. For some methods integer and logic, rather than floating point, operations will expedite building **G** many-fold.

INTRODUCTION

Genomic data can provide more accurate information about relationships among animals. When only pedigree information is available, progeny are assumed to receive a random half of each parents' genes and full-sibs are expected to share half their genes. With genomic data we can tell which half of each parents' genes an animal receives and precisely the proportion of genes shared by full-sibs. Generally, genomic information provides more detailed information about relationships including that between individuals that share no known common ancestors.

When a population is genotyped a genomic relationship matrix (**G**) takes the place of the numerator relationship matrix (**A**) in routine genetic analyses. However, unlike **A**, **G** must be built explicitly which can be a time consuming process particularly when the number of loci and/or genotyped animals is large. When $\mathbf{G}^{-1}$ is needed, **G** must also be inverted directly as it is dense and unlike **A**, **G** has no simple inverse. This operation is generally more computationally expensive than building **G** whereas $\mathbf{A}^{-1}$ can be constructed rapidly, directly from the pedigree.

Recently Forni *et al.* (2012) examined the effect of using different assumptions to build **G** but obtained the same results for some methods. This paper illustrates how using different assumptions when building **G**, can result in different **G** matrices and even estimated genetic variances, yet provide the same estimated breeding values (EBVs). It also shows how different assumptions can significantly expedite the process of building **G**.

THEORY

Estimates of relationships among individuals are essential for genetic evaluation. Traditionally **A** fulfilled that purpose. When combined with the genetic variance (σ_u^2), variance of the breeding values (**u**) was defined to be $\mathrm{Var}(\mathbf{u})=\mathbf{A}\sigma_u^2$. **A** is based on the idea of identity by descent (IBD) and is built by tracing the flow of genes down the pedigree. Elements of **A** are twice the coancestry coefficient which are probabilities that limit the range of elements in **A** to [0,2]. Founders, the remotest set of ancestors with unknown pedigree, are assumed to be a random sample from a very large population in Hardy-Weinberg equilibrium. The partition of **A** relating to the founders is an identity matrix, which implies that the genome of each founder consists of two subsets. The first subset contains loci that are all homozygous and common to all founders and thus generate no

* AGBU is a joint unit of NSW Primary Industry and the University of New England

phenotypic variance. The other subset contains all loci that generate phenotypic variation. They are unique to each founder as off-diagonal elements of zero imply that there is no covariation with any other founder. This suggests that there were an infinite number of alleles at every locus in the base population.

Genomic data, in the form of single nucleotide polymorphisms (SNP), can be used to build **G** (Van Raden 2008) for individuals with genotypes. Using markers involves the strong assumptions relating to identity by state (IBS), where markers are deemed to be in linkage disequilibrium with genes affecting phenotypes, and that such genes behave similarly across the whole population, especially for relationships beyond the pedigree. When all individuals in the population have genotypes then **G** can be used in place of **A** so that the assumption about the variance of the breeding values becomes $\text{Var}(\mathbf{u})=\mathbf{G}\sigma_u^2$. A variety of different methods are available for building **G** and some of them are equivalent to including the SNP directly as individual effects (**g**) in the model (Stranden and Garrick, 2009) in place of the breeding values, so that $\mathbf{u}=\mathbf{Zg}$ and $\text{Var}(\mathbf{g})=\mathbf{I}\sigma_g^2$, where σ_g^2 is the variance due to the SNPs. The equivalence between these methods indicates a degree of ambiguity and loosely implies that the effects of the SNPs, or the quantitative trait loci in linkage disequilibrium with them are estimable. Some methods for building **G** result in elements that have no probabilistic interpretation (e.g. elements less than zero).

Genomic data. SNPs are the genotypes used in this paper, with each individual-locus represented by a number 0, 1 or 2, being the number of one of the alleles available at the locus. There are a animals with h haplotypes (h=2a) and m loci. The genotypes are represented by **Z**, an a x m matrix and haplotypes by **X** an h x m matrix. Haplotypes for each locus are formed independently of other loci. The matrix $\mathbf{K}=\mathbf{I} \otimes [1\ 1]$, where $\otimes$ is the Kronecker product, converts **X** to **Z** as $\mathbf{Z}=\mathbf{KX}$. The matrix **P** is conformable to **Z** and contains the allele frequencies (p) for each locus in its columns. In addition let **J** denote a matrix with all elements equal to 1. Dimensions of **J** are as implied in the equation where it is used. Where necessary we specify the row (i) and column (j) dimensions as subscripts ($\mathbf{J}_{ij}$).

G matrices. Three alternative methods for building **G** are considered. The first of these is Van Raden's (2008) first method, viz. $\mathbf{G}=\mathbf{MM}'/d$, where $\mathbf{M}=\mathbf{Z}-2\mathbf{P}$, and $d=2\sum p(1-p)$. By subtracting 2**P** from **Z** genotypes are centred so that columns of **M** sum to zero. The denominator is designed to scale the matrix **G** to be similar to the scale of **A**. This formulation of **G** generates some irregular elements that cannot be interpreted as co-ancestry. These include negative elements, parent-offspring elements less than 0.5 and diagonals less than 1. Potentially, elements can be greater than 2 (between pairs of individuals sharing a very large number of low frequency alleles).

The second method is similar to the first with genotypes centred around zero: $\mathbf{F}=(\mathbf{Z}-\mathbf{J})(\mathbf{Z}-\mathbf{J})'/c$. The denominator, c, can be the same as d, or alternatively with all allele frequencies set to 0.5, c=m/2. **F** can also contain unusual elements, with the diagonal elements being a function of the proportion of the animals' loci that are homozygous. Elements of **F** are readily computed by counting the numbers of identical and of opposing homozygotes between each pair of animals. This allows the use of integer and logical operations that are much faster than floating point operations required to compute $(\mathbf{Z}-2\mathbf{P})(\mathbf{Z}-2\mathbf{P})'$.

The third method is based on building a gametic relationship matrix (**H**). Nominally, a gametic relationship matrix ($\mathbf{\Gamma}_i$) is built for each locus by counting 1 if the alleles are the same and 0 if they differ. Subsequently the complete gametic relationship matrix ($\mathbf{\Gamma}$) is calculated by summing all the loci matrices and dividing by m. This is converted to the animal relationship as $\mathbf{H} = \mathbf{K\Gamma K}'/2$. In practice, **H** is built as $\mathbf{H} = \mathbf{K}[\mathbf{XX}'+(\mathbf{X}-\mathbf{J})(\mathbf{X}-\mathbf{J})']\mathbf{K}'/2m$. The method for building **H** ensures that it has no elements less than 0 nor greater than 2 and no diagonal elements less than 1.

Similarity. Expansion of the terms in the matrices illustrates the differences between them.

1. Considering **M** as $\mathbf{Z}-\mathbf{J}-\mathbf{D}$, where $\mathbf{D}=2\mathbf{P}-\mathbf{J}$ the numerator of **G** ($=\mathbf{MM}'/d$) gives

$\mathbf{MM'} = (\mathbf{Z}-\mathbf{J}-\mathbf{D})(\mathbf{Z}-\mathbf{J}-\mathbf{D})' = (\mathbf{ZZ'}-\mathbf{ZJ'}-\mathbf{ZD'}-\mathbf{JZ'}+\mathbf{JJ'}+\mathbf{JD'}-\mathbf{DZ'}+\mathbf{DJ'}+\mathbf{DD'})$

By setting $\mathbf{E} = -\mathbf{ZD'}+\mathbf{JD'}-\mathbf{DZ'}+\mathbf{DJ'}+\mathbf{DD'}$ and noting that with $\mathbf{J}_{am}\mathbf{J}_{am}'=m\mathbf{J}_{aa}$,

$\mathbf{G} = (\mathbf{ZZ'}+m\mathbf{J}-\mathbf{ZJ'}-\mathbf{JZ'}+\mathbf{E})/d$

2. $\mathbf{F} = (\mathbf{Z}-\mathbf{J})(\mathbf{Z}-\mathbf{J})'/c = (\mathbf{ZZ'}+m\mathbf{J}-\mathbf{ZJ'}-\mathbf{JZ'})/c$.
3. $\mathbf{H} = (\mathbf{K}[\mathbf{XX'}+(\mathbf{X}-\mathbf{J})(\mathbf{X}-\mathbf{J})']\mathbf{K}/2)/m$

   $= (\mathbf{KXX'K'}+\mathbf{KJJ'K'}/2-\mathbf{KXJ'K'}/2-\mathbf{KJX'K'}/2)/m$, and since $\mathbf{Z} = \mathbf{KX}$ and $\mathbf{KJ}_{hm} = 2\mathbf{J}_{am}$,

   $\mathbf{H} = (\mathbf{ZZ'}+2m\mathbf{J}-\mathbf{ZJ'}-\mathbf{JZ'})/m$.

These results clearly show how **G**, **F** and **H** differ and that since, $\mathbf{G} = (\mathbf{F}c+\mathbf{E})/d$ and $\mathbf{F} = m(\mathbf{H}-\mathbf{J})/c$, how one can be determined from another. When c=d, $\mathbf{G} = \mathbf{F}+\mathbf{E}/d$.

MATERIALS AND METHODS

A small population made up of four sires mated to the same five dams each producing one offspring was generated. Each individual had two haplotypes of 99 SNPs, a breeding value and phenotype for a trait with a heritability of 0.55. These were analysed with the model $\mathbf{y}=\mu+\mathbf{Z}_1\mathbf{u}+\mathbf{e}$, where the data are a function of the mean (μ), the breeding values (**u**) and a residual (**e**), and $\mathbf{Z}_1$ is an incidence matrix assigning observations to breeding values. $\text{Var}(\mathbf{u})=\mathbf{W}\sigma_u^2$, where **W** is a relationship matrix and $\text{Var}(\mathbf{e})=\mathbf{I}\sigma_e^2$. Genetic parameters for this population were estimated using five different matrices **W.** The first used **G** with a small amount (0.01**I**) added to make it invertible (positive definite), the second and third used **F** with denominators of d and m/2 respectively, the fourth used **H** and the last used **F**+10**J**. These data were analysed with WOMBAT (Meyer, 2007) to estimate variance components and breeding values.

RESULTS AND DISCUSSION

G matrices. The construction of the various matrices shows clearly how they relate to each other. The difference between **G** and **F**(c=m/2) arises from the different allele frequencies. **F** and **H** differ in their denominators and there is an additional term (m**J**) included in **H** that is not in **F**.

Evaluations. The results in Table 1 show that, regardless of which **W** matrix is used, the estimated breeding values (EBVs) are the same. The correlations between EBVs from different analyses are 1, or close to 1, as are the regressions of 1 on those obtained when **W**=**G**. Differences in estimated means are unimportant as EBVs are relative measures of genetic merit. Slight differences occur when **G** is used, compared to the other methods as its diagonal was augmented and some difference in the mean may be due to **E**. The addition of 10**J** to **F** has no effect, indicating that adding any multiple of **J** (results not shown) to these matrices have no effect. These results show the practice of augmenting the diagonal of **G** should be superseded by adding k**J**, where k is small, to ensure **G** is invertible. The likelihoods and residual variances are also the same for all models. Similar genetic variances were estimated when **G** or **F** was used. While the addition of a multiple of **J** to **F** matrices has no effect, it suggests a higher degree of relationship in that population than **F** alone. Using **H** obtained a considerably higher additive genetic variance

Table 1: **Results from evaluation of simulated data using different relationship matrices**

Relationship Matrix	Log-Likelihood	σ_e^2	σ_u^2	μ	Regression of EBVs on EBVs(**G**) Intercept	Slope	Correlation EBVs with EBV(**G**)
G	-76.55	51.22	31.74	0.000	-	-	-
F(c=d)	-76.55	51.27	31.71	-0.027	0.027	0.999	1.0
F(c=m/2)	-76.55	51.27	33.83	-0.027	0.027	0.999	1.0
H	-76.55	51.27	67.73	-0.027	0.027	0.999	1.0
F+10**J**	-76.55	51.25	31.74	-0.027	0.027	0.997	1.0

than other matrices. This might suggest that **H** uses a more ancient set of founders than assumed when **G** or **F** is used. However, since Var(**u**)=$\mathbf{W}\sigma_u^2$, and if it is only their denominators that differ ($\mathbf{W}_1$=w$\mathbf{W}_2$), the estimated additive genetic variance must vary in a complementary manner ($\sigma_{u1}^2=\sigma_{u2}^2/w$). This is so for **F**(c=d) and **H** where the ratio of the additive genetic variances is d/m and similarly for **F**(c=d) and **F**(c=m/2) where this ratio is 2c/m.

Although the various genomic relationship matrices were different, their inverses, also necessarily different, provide the same results which may seem surprising given the different assumptions. Despite this, the same results indicate that the inverses are simple functions of each other showing that the genomic data are being used in exactly the same way.

The equivalence between these methods, based on relationship matrices, can be illustrated by considering modelling the genotypes directly. With this model the addition of a constant to the SNP genotypes for each locus has no effect on anything but the overall mean. The additive breeding values (**u**=**Zg**) would be the same as if nothing had been added. This is akin to centering alleles around different values and adding terms like **E** and k**J** to any **W**.

These results show that different approaches to using genomic data may not ensure real differences and may explain why some methods used by Forni *et al.* (2011) have identical results. These results also show that the apparent problems relating to strange elements (negative off-diagonals, and diagonals less than 1) in **G** are nothing to fear, they are simply on a different scale to the other **W**s. Starting with the idea of SNP similarity provides **H** which, by construction, can have a similar probabilistic interpretation to **A**. However, **H** provides a much greater genetic variance than the other methods, but this can be modified by factoring it by c/m.

As genomic data provide relationships among individuals that are not IBD, it is clear that the unknown founder population implied when genomic data are used must be different to the known founder population derived from pedigrees. These results show that the estimated additive genetic variance is sensitive to assumptions about allele frequencies which determine the denominator and, indirectly, the unknown founder population. Paradoxically, the EBVs estimated from each of these evaluations are insensitive to the different estimates of additive genetic variance when combined with the appropriate **W**. Conversely, incorrect EBVs could result from combining a relationship matrix **W** with an inappropriate additive genetic variance.

Building the numerators of **F** and **H** are based on **Z** and **X**. These matrices are integers and provide the opportunity to use integer rather than floating point operations. Furthermore, as the non-zero elements of **Z**-**J** are only 1, and -1 the process of building **F** can be done with logic operators which is magnitudes faster than the floating point operations used to build **G.**

CONCLUSION

Many ways of using genomic data to determine relationships among individuals in a population, while appearing to be different, are similar. Although they may be based on different assumptions, and can provide different estimates of the additive genetic variance, they provide the same measures of genetic merit of the population. The estimate of the additive genetic variance is sensitive to the estimate of allele frequencies. **F** should be used in place of **G,** as it is much quicker to build and provides an equivalent model and it does not require augmenting the diagonal to make it invertible.

REFERENCES

Forni, S., Aguilar, I., and Misztal I. (2011) *Genet. Sel. Evol*. **43**:1.
Meyer, K. (2007) *J. Zhejiang Univ. Sci. B* **8**:815.
Stranden I., Garrick D. (2009) *J. Dairy Sci*. **92**:2971.
Van Raden P.M. (2008) *J. Dairy Sci*. **91**:4414.

Proc. Assoc. Advmt. Breed. Genet. 21: 465-468

DETECTION AND VALIDATION OF STRUCTURAL VARIATION IN CATTLE WHOLE-GENOME SEQUENCE DATA

L. Chen[1, 2, 3], A.J. Chamberlain[1,3], C.M. Reich[1, 3], H.D. Daetwyler[1, 2, 3] and B.J. Hayes[1, 2, 3]

[1]AgriBio, Centre for AgriBioscience, Biosciences Research, DEDJTR, 5 Ring Road, Bundoora, VIC 3083, Australia
[2] La Trobe University, Victoria, Australia
[3] Dairy Futures Cooperative Research Centre, Victoria, Australia

SUMMARY

Several examples of structural variation (SV), or copy number variation (CNV) affecting traits exist in cattle. However the effect of SV on complex traits is largely unknown. The identification of SV suffers from high false positive and low overlapping rate when using different programs. We detected SV in dairy cattle whole-genome sequence from 308 Holsteins and 64 Jerseys with two SV detection programs - Breakdancer and Pindel. We constructed a set of validated SVs based on 28 individuals that were sequenced twice, and were transmitted sire to son. A total of 11,534candidate SVs covering 5.64 Mb were validated in the 28 twice-sequenced individuals, while 3.49 Mb and 0.67 Mb of SV were validated from Holstein and Jersey sire-son transmission.

INTRODUCTION

There are several categories of genome variation within a species. Single nucleotide polymorphisms (SNP) are the most frequent and have been widely utilized in association and genomic prediction studies. Another category is structural variation (SV) which refers to segments of 1 kilo bases (kb) to several mega bases (Mb) of deletions, duplications, inversions and translocations in the re-sequenced genome compared to a reference genome, of which copy number variation (CNV) only includes deletions and duplications.

In cattle, a number of studies have shown evidence that SVs spanning gene coding regions can affect a wide range of traits (Liu *et al.* 2010). In Angus cattle, 297 CNVs were found to be associated with parasite resistance or susceptibility (Hou *et al.* 2012). Recently a 660 kb deletion was found to be associated with fertility and milk production in Nordic red cattle (Kadri *et al.* 2014). In addition, SVs have been shown to be associated with the polled phenotype in cattle (Medugorac *et al.* 2012; Rothammer *et al.* 2014)

A number of genomic data types can be used to detect SV. PennCNV implements a hidden Markov model (HMM) to detect CNVs from SNP arrays (Wang *et al.* 2007). However, due to limited SNP density and high minor allele frequency of these SNP, the ability to identify rarer and/or smaller CNVs is limited. In addition, SNP chip methods cannot capture balanced SVs including inversions and translocations.

Whole-genome sequence data can potentially be used to recover the whole spectrum of SVs. Paired-end mapping (PEM) (Korbel *et al.* 2007), split read (SR) (Ye *et al.* 2009), read depth (RD) (Teo *et al.* 2012), and de novo assembly (Iqbal *et al.* 2012) are the current four basic strategies used to detect SVs from sequence data.

Here we detected SVs in whole-genome sequence data from Holstein and Jersey populations with a combination of Breakdancer (Chen *et al.* 2009) (PEM) and Pindel (Ye *et al.* 2009) (SR), combined with two novel validation strategies, to generate high quality SV sets. We also tested the hypothesis that highly conserved gene regions (between species) should have less SVs than in less conserved regions.

MATERIALS AND METHODS

Animal samples. The paired-end read whole-genome sequence data is described in (Daetwyler *et al.* 2014). A total of 308 Holstein and 64 Jersey were sequenced with Illumina sequencing platforms, with average coverage 10.76 and 10.92 respectively. All the short sequencing reads were then aligned to reference assembly UMD 3.1 with the Burrows-Wheeler Aligner (BWA). Our validation strategy included assessing how many SVs were detected in both replicates of a set of 28 Holstein individuals that were sequenced twice with different libraries, and whether we could observe sire-son transmission of the SV in 68 Holstein and 33 Jersey sire-son pairs.

Sequence population SV calls. We pooled the Holstein (not including twice-sequenced individuals) and Jersey populations and investigated the SV distribution differences between the two breeds. For each population, we first ran Breakdancer and Pindel to generate raw SV calls by each SV type (deletion, insertion, inversion and duplication). The default parameters were used for both programs. However, we enforced a threshold of a minimum of four supporting read pairs and observation in two individuals to classify higher quality SVs. We also filtered SVs that span chromosome gaps in the reference assembly. In the next step, we found the overlapping regions when merging the SV calls from Breakdancer and Pindel and considered these overlapped regions to be higher confidence SVs.

Validated SV calls. In the Holstein population, 28 individuals were sequenced twice. In theory for each individual the two sequences should convey exactly the same information. However due to random distribution of sequence reads, assembly error and different depth of coverage, the two sequences are not identical, and, thus, programs can report different SVs. We generated a high confidence SV set by only reporting SVs detected in both sequences. In addition, as most SVs should be inherited, we only report SVs that are inherited from sire to offspring. The validated sets were further compared between each other and with outputs from SNP chip.

Detecting SVs and CNVs from SNP chip genotype data. A total of 128 Holstein and 170 Jersey cattle were genotyped with the 800K HD SNP chip, which were afterwards converted to Log R Ratio (LRR) and B allele frequency (BAF) for further analysis. Individuals with standard deviation of LRR>0.35 and BAF >0.2 were discarded, as suggested by Wang *et al.* 2007. A total of 125 Holstein and 166 Jersey were kept after this filter. The genomic content (GC) model which incorporates the GC percentage information around each SNP was used to improve CNV outputs. SNP chip methods cannot detect inversions and therefore we eliminated inversion events when comparing to validated sets from sequence.

Conserved genes. To test the hypothesis that SV and CNV are less likely in genes that are highly conserved across species, 248 core eukaryotic genes were selected (Parra *et al.* 2007) that were likely to be found in a low number of paralogs in a wide range of species. We downloaded the protein file (fasta format) and put it into the BLAST program to search the most similar proteins and genes in cattle. The search results were further converted into coding nucleotides in bed format with chromosome, strand, start and end position that can be overlapped with our validated SV sets. We defined a minimum of 0.5% of the gene overlapped with validated SVs to be reported. A chi-squared test was performed to test whether these conserved genes contain less SVs than all the other reference genes downloaded from the UCSC genome browser.

RESULTS AND DISCUSSION

Population SV Calls. The overlapped region from the two programs dramatically shrunk the original SVs into a small set, as only about 2-10% of the calls (ranging from 25 to 44,412 bp) were

kept after merging (p-value = 6.38448e-20). Overall, Holstein had more SV calls than Jersey, which may mainly be due to a larger sample size for Holstein. After filtering SVs less than 25 bp, the median length of deletion, insertion, inversion and duplication for Holstein was 1123, 72, 2533 and 857 and for Jersey was 1152, 0, 1337 and 1014 bp, respectively. Table 1 shows the total covered length of SVs shared by the two populations. A total of 4.62Mb SV events were detected in both population, occupying 16.89% in Holstein (27.36 Mb) and 53.47% in Jersey (8.64 Mb), of which deletions and duplications had a relatively high percentage.

Table 1. Covered region of SVs shared by Holstein and Jersey population

SV Covered Region Mb	DEL	INS	INV	DUP	Total
Holstein	8.49	0.639	13.84	4.40	27.36
Jersey	5.22	0	1.05	2.37	8.64
OVERLAP	3.18	0	0.22	1.23	4.62

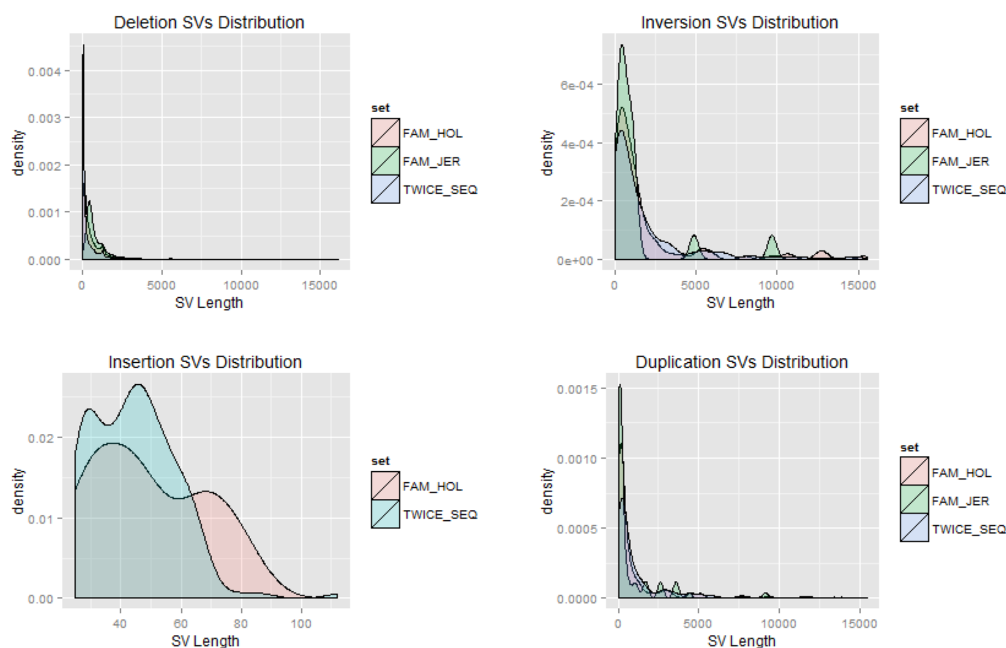


Figure 1. Size range distribution of four type of SVs in twice sequenced, Holstein and Jersey family validated sets.

Validated SV Calls. We generated three sets of validated SV calls: twice-sequenced, Holstein and Jersey family-level validated SV sets. A total of 5.64 Mb were validated from 28 twice-sequenced individuals, while 3.49 Mb and 0.67 Mb SVs were found in Holstein and Jersey families. We also compared the Holstein twice-sequenced set and Holstein family set. Overall 82.0% SVs in Holstein family were also found in the twice-sequenced set. This result illustrates less false positives and thus higher confidence SVs compared to population calls. Figure 1 demonstrates that the size distribution of SVs is similar across these validated sets. Most deletions and insertions are

less than 100 bp; a large number of inversions are around 900 bp while duplications are around 350 bp. For inversions in Jersey family there are two small peaks at 5kb and 10 kb respectively. When looking into the sires with multiple sons, a total of about 80 kb deletions and 90 kb duplications on BTA1 were shared in Holstein and 27 kb inversions on BTA11 and 16 kb inversions and duplications on BTA14 in Jersey, suggesting these areas could be common CNV regions in both breeds.

The 800K SNP chip data results indicated a total of 2224 CNVs covering 250.5 Mb in Holstein (227 Mb deletions and 23.3 Mb insertions) and 2976 CNVs covering 357.4 Mb in Jersey (333 Mb deletions and 24.3 Mb insertions). As SNP platform resolution is limited, PennCNV cannot detect very small events. Therefore, we only compared this result with SVs larger than 5 kb detected from the sequence data. As a result, 12.33% deletions and 11.59% duplications in validated sets were also found in Holstein 800K outputs, while 14.95% deletions and 0% insertions overlapped in Jersey.

Conserved Genes Test. We found 293 identical genes according to core gene sets after searching by BLAST. Overall, there were not many conserved genes in our reported SV areas. Within the 293 genes only five genes were found in Holstein family, one in Jersey (*ETFDH* with 152 bps overlapped) and seven in twice-sequenced one. Among these genes, most harboured deletions, while two and one contained inversions and a duplication, respectively. All the five genes from the Holstein family set were confirmed in the twice-sequenced set. Compared to all the other reference sequence genes, however, no significant evidence was found to support that conserved genes regions contained less structural variants than all others (p-value >0.7). Our validated SV sets will assist genetic research in cattle such as genomic prediction and genome-wide association studies.

ACKNOWLEDGEMENTS

The authors thank all members of the 1000 Bull Genomes Consortium for provision of data.

REFERENCES

Chen K., Wallis J.W., McLellan M.D., Larson D.E., Kalicki J.M., *et al.*(2009) *Nat Methods* **6**(9): 677.

Daetwyler H.D., Capitan A., Pausch H., Stothard P., van Binsbergen R., *et al.* (2014) *Nat Genet* **46**(8): 858.

Hou Y., Liu G.E., Bickhart D.M., Matukumalli L.K., Li C., *et al.* (2012) *Funct Integr Genomics* **12**(1): 81.

Iqbal Z., Caccamo M., Turner I., Flicek P., McVean G., (2012) *Nat Genet* **44**(2): 226.

Kadri N.K., Sahana G., Charlier C., Iso-Touru T., Guldbrandtsen B., *et al.* (2014). *PLoS Genet* **10**(1): e1004049.

Korbel J.O., Urban A.E., Affourtit J.P., Godwin B., Grubert F., *et al.* (2007) *Science* **318**(5849): 420.

Liu G. E., Hou Y., Zhu B., Cardone M.F., Jiang L., *et al.* (2010) *Genome Res* **20**(5): 693.

Medugorac I., Seichter D., Graf A., Ingolf R., Helmut B., *et al.* (2012) *PLoS One* **7**(6): e39477.

Parra G., Bradnam K. and Korf I. (2007) *Bioinformatics* 23(9):1061

Rothammer S., Capitan A., Mullaart E., Seichter D., Russ I., *et al.*(2014) *Genet Sel Evol* **46**: 44.

Teo S. M., Pawitan Y., Ku C.S., Chia K.S. and Salim A. (2012) *Bioinformatics* **28**(21): 2711.

Wang K., Li M., Hadley D., Liu R., Glessner J., *et al.* (2007) *Genome Res* **17**(11): 1665.

Ye K., Schulz M. H., Long Q., Apweiler R. and Ning Z.M. (2009) *Bioinformatics* **25**(21): 2865.

Proc. Assoc. Advmt. Breed. Genet. 21: 469-472

A COMPARISON OF GENETIC CONNECTEDNESS MEASURES USING DATA FROM THE NZ SHEEP INDUSTRY

J.B. Holmes[1], B. Auvray[1], S.A. Newman[2], K.G. Dodds[2] and M.A. Lee[1]

[1] Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand
[2] AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

SUMMARY
The New Zealand Sheep industry, via Sheep Improvement Limited (SIL), estimates genetic connectedness across flocks as a function of progeny counts. This estimate is derived separately from the model fitted to estimate breeding values. As it ignores sources of genetic linkages other than direct parent-progeny links, it may under-estimate the level of connectedness present in the flocks assessed. In this paper, we compared this estimate to another derived from the variance-covariance (relationship) matrix of additive effects when pedigree information was available and when genotype information was available on some of the animals assessed. For the example of a single trait model using weaning weight records, we found an increase in the level of connectedness estimated compared to the existing method, particularly when genotype information was incorporated in the relationship matrix.

INTRODUCTION
To optimise genetic gain in livestock programs, breeding values need to be predictable between flocks. In animal breeding literature this is referred to as connectedness. In the BLUP methods used to estimate breeding values, the most appropriate measure of connectedness is the prediction error variance-covariance matrix (PEV). However, this calculation is computationally demanding and many proxies have been proposed.

The standard error of differences in breeding value means between flocks can be estimated as a function of the number of progeny born to common parents across flocks. This approximation is often used in traditional evaluations, where only pedigree information is used, but it is problematic when genotype information is also incorporated. As genetic evaluations for New Zealand sheep are increasingly using genotype data, a measure of connectedness derived from the model is preferred so that we can quantify genetic connectedness that is due to including genotype data.

In this paper we compared the standard error of differences in breeding value means calculated from a model based proxy to PEV, (i.e. genetic drift variance (Kennedy and Trus 1993)) to the current measure. This was done for scenarios where only pedigree information was available, and when some animals had genotype information available.

MATERIALS AND METHODS
Data. The data was from 64,841 animals from 19 flocks born from 2011 to 2013 with weaning weight records. The pedigree file containing the recorded animals and parents without records consisted of 84,802 animals. Genotype information (50K Illumina SNP Chip) was available for 269 of these animals of which 21 were in the initial set with weaning weight records. There were 31,884 animals that were either genotyped or had a genotyped ancestor. Table 1 shows the distribution of animals with genotype records or a genotyped ancestor across flocks.

Table 1. Distribution of animals with weaning weight records and genotype records on either themselves or at least one parent across flocks

Flock	Number with records	Number with a genotyped ancestor	Percentage with genotypes.
1	641	0	0.00
2	2533	2065	81.52
3	21240	14404	67.82
4	1996	1314	65.83
5	2344	1513	64.55
6	1110	0	0.00
7	16761	8231	49.11
8	1984	1785	89.97
9	815	769	94.36
10	3535	0	0.00
11	787	0	0.00
12	953	0	0.00
13	1025	699	68.20
14	2412	293	12.15
15	1193	528	44.26
16	368	0	0.00
17	2226	222	9.97
18	984	0	0.00
19	1934	61	3.15

SIL measure. The measure of connectedness between two flocks used for genetic evaluations performed in SIL is proportional to the standard error of the weighted average of differences of breeding values (u) between flocks across parents namely

$$\sqrt{\sum_i \lambda_i^2\left(\frac{1}{n_{A_i}}+\frac{1}{n_{B_i}}\right)} = \sqrt{1/\sum_j\left(\frac{1}{n_{A_j}}+\frac{1}{n_{B_j}}\right)^{-1}} \qquad [1]$$

where n_{A_j} is the number of progeny of parent *j* in flock A, n_{B_j} is the number of progeny of parent *j* in flock B and $\lambda_i \propto \left(\frac{1}{n_{A_i}}+\frac{1}{n_{B_i}}\right)^{-1} \Big/ \sum_j\left(\frac{1}{n_{A_j}}+\frac{1}{n_{B_j}}\right)^{-1}$. The standard error of differences has a range of $(0,\sqrt{2}]$. If there are no progeny from common parents in flock A and flock B, the standard error of the difference was arbitrarily set to 2. Only progeny born in a set time period are considered when calculating this measure. This is usually taken to be the previous three years, and has also been applied in this paper.

Variance-covariance matrix measure. The standard error of differences in average breeding values between flock A and B was calculated to proportionality from the elements of $\boldsymbol{V} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'ZGZ'X}(\boldsymbol{X'X})^{-1}$ (Kennedy and Trus 1993) corresponding to flock A and B.

$$S.E.(\bar{u}_A - \bar{u}_B) \propto \sqrt{\boldsymbol{V_{AA}} + \boldsymbol{V_{BB}} - 2\boldsymbol{V_{AB}}} \qquad [2]$$

where $\boldsymbol{G}$ is an additive relationship matrix, $\boldsymbol{Z}$ is the incidence matrix of animals with records and $\boldsymbol{X}$ is the flock incidence matrix. Two formulations for $\boldsymbol{G}$ were used. When only pedigree information was available, which we refer to as the pedigree measure, $\boldsymbol{G} = \boldsymbol{A}$, the pedigree additive relationship matrix. When some animals had genotype information available, which we refer to as the single step measure $\boldsymbol{G} = \boldsymbol{H}$. To calculate $\boldsymbol{H}$ both $\boldsymbol{A}$ and a genomic relationship matrix $\boldsymbol{G_1}$ was

required. $\boldsymbol{G_1}$ was calculated for the genotyped animals using the first method of VanRaden (2008). The $\boldsymbol{H}$ matrix was constructed using the method in Aguilar et al. (2010), where $\boldsymbol{A_{11}}$ and $\boldsymbol{A_{22}}$ are the additive relationship matrices for, and $\boldsymbol{A_{12}}$ is the matrix of additive relationship covariances between the un-genotyped and genotyped animals respectively.

$$\boldsymbol{H} = \boldsymbol{A} + \begin{bmatrix} \boldsymbol{A_{12}A_{22}^{-1}(G_1 - A_{22})A_{22}^{-1}A_{21}} & \boldsymbol{A_{12}A_{22}^{-1}(G_1 - A_{22})} \\ \boldsymbol{(G_1 - A_{22})A_{22}^{-1}A_{21}} & \boldsymbol{G_1 - A_{22}} \end{bmatrix} \quad [3]$$

When $\boldsymbol{V_{AB}} = 0$, the standard error of the difference in average breeding value is set to 2, analogously to the situation of no progeny from common parents in the SIL measure.

RESULTS AND DISCUSSION

Clusters of connected flocks. The connectedness estimated from the different methods is given in Table 2 where clusters of flocks estimated to be connected are shown. The criterion to be connected is a standard error of difference less than 2. The measure currently used in the NZ genetic evaluation was the most conservative in estimating connectedness across flocks and the single step measure was the least conservative. Changes in the clustering between the three measures were due to the admission of previously isolated flocks into clusters, or cluster merging rather the shifting of flocks from one cluster to another. This made intuitive sense since any linkage coming from shared parents is also contained in the pedigree along with linkage from more distant ancestors, such as grandparents. In turn in the genomic relationship matrix, almost all off-diagonals are non-zero, even for animals thought to be unrelated.

Table 2. Clusters of linked flocks (identified by flock code) according to the three measures of connectedness used

				SIL measure					
Cluster 1	2	8	13						
Cluster 2	3	7	17						
Isolated Flocks	1	4	5	6	9	10	11	12	14
	15	16	18	19					
				Pedigree measure					
Cluster 1	2	4	8	13					
Cluster 2	3	7	14	17					
Cluster 3	1	12	19						
Cluster 4	5	16							
Cluster 5	6	10							
Cluster 6	11	18							
Isolated Flocks	9	15							
				Single step measure					
Cluster 1	1	2	3	4	5	7	8	9	12
	13	14	15	16	17	19			
Cluster 2	6	10							
Cluster 3	11	18							

Comparison of standard error of differences. Figure 1 plots the standard error of differences for the three measures considered. A reduction in the number of standard errors being arbitrarily set to 2 was found moving from the SIL measure to the pedigree and single step measures. This corresponded to the reduction in isolated flocks found in the cluster analysis. For flock pairs where connections were found using the pedigree measure, the rank correlation of the standard error of

differences between the pedigree and single step measure was 0.9902.

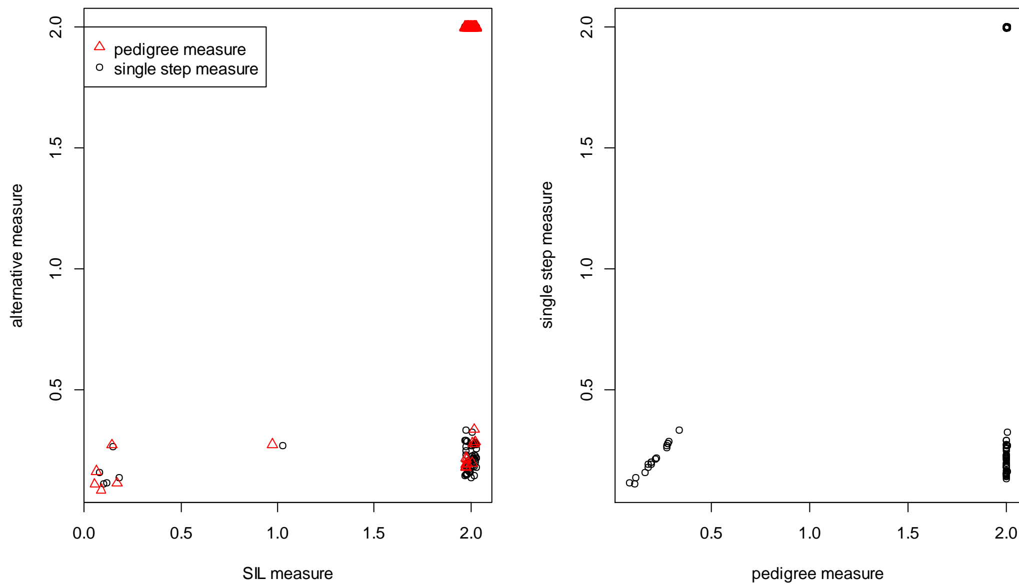


Figure 1. Comparison of standard error of differences using the three measures.

The loss of connections through removed data. In this paper, we used three years of data in the calculation of the connectedness measure. In routine genetic evaluations, there are many more years of records and pedigree data available. It may be inappropriate to develop a flock based connectedness measure from the full relationship matrix from a routine genetic evaluation, since connections from old animals would be given equal weighting to younger animals. Kennedy and Trus (1993) discussed changing the incidence matrix $\boldsymbol{X}$ in genetic drift variance from flock to flock by year. This method would utilise the connections lost through data removal while removing bias in measured connectedness through equal weighting of older and younger animals.

Single step method results. The single step measure assigned all flocks with genotype information and any flock related to such a flock through pedigree into a single cluster of related flocks. This means single step BLUP would lead to an increase in the number of animals with comparable breeding values compared to traditional BLUP. It is unclear how reasonable this result is, but this warrants further investigation. However the genomic relationship matrix used uses IBS to estimate relatedness. As a result the accuracy and comparability of estimated breeding values may be inflated. The degree of this inflation will be dependent on the group of markers used in the calculation of the genomic relationship matrix and the ancestry of the animals. To overcome this, developing an unbiased estimation of a genomic relationship matrix may be of interest.

REFERENCES

Kennedy B.W., and Trus D. (1993) *J. Dairy Sci.* **71**: 2341.
VanRaden P.M. (2008) *J. Dairy Sci.* **91**: 4414.
Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsurata S. and Lawlor T.J. (2010) *J. Dairy Sci.* **93**: 743

Proc. Assoc. Advmt. Breed. Genet. 21: 473-476

EVOLVING TO THE BEST SNP PANEL FOR HANWOO BREED PROPORTION ESTIMATES

C. Esquivelzeta-Rabell[1], H.A. Al-Mamun[1], S.H. Lee[2], H.K. Lee[3], K.D. Song[3] and C. Gondro[1]

[1]School of Environmental and Rural Science, University of New England, NSW, Australia
[2]Division of Animal and Dairy Science, Chung Nam National University, Daejeon, Korea
[3]The Animal Genomics and Breeding Center, Hankyong National University, Anseong, Korea

SUMMARY

Hanwoo is highly prized for its marbling ability and is the most important cattle breed in Korea. In order to maintain the integrity of the breed and for product certification purposes it is important to develop tools to confirm the origin of the products. Breed composition estimates based on a large number of molecular markers (e.g. HD SNP arrays) are highly accurate but expensive for routine usage. The identification of a reliable panel with a small number of markers will reduce costs and can enable broader adoption of the technology by industry. In this work a heuristic optimization method was used to find the most reliable subset of markers, from the Illumina BovineHD array, to estimate breed proportion in Hanwoo. Accuracies of breed proportion estimates above 90% can be achieved using as little as 200 markers. The best balance between accuracy and number of SNP was obtained with 500 markers achieving 94% accuracy. Rapid and cost effective breed composition prediction in Hanwoo cattle based on a SNP panel with at least 200 markers will help to certify the products with an acceptable accuracy and ensure breed purity within the breeding program. The method described herein is directly applicable to other breeds.

INTRODUCTION

Hanwoo is the most important native Korean cattle and its history traces back 5,000 years (Jo *et al.* 2012). Over this long timespan the purpose of these cattle has evolved from farming, transportation and religious sacrifice to beef production (Lee *et al.*, 2014). Hanwoo beef has unique marbling characteristics which confer a special tenderness, juiciness and unique flavour to the meat, making it highly sought after by consumers at premium prices (Kim *et al.* 2000; Han and Lee 2010; Jo *et al.* 2012). It has also been shown that Hanwoo has a healthier fatty acid composition in comparison to other breeds (Jo *et al.*, 2012) which makes them even more attractive to consumers. In order to certify the products it is important to develop cost effective tools that allow verifying that the product truly comes from pure bred Hanwoo cattle. Breed prediction is also a useful tool for breed associations where the animals need to be purebred to be registered and, within genomic selection (GS) programmes, it can be used for quality control of research and industry samples (Dodds *et al.*, 2014).

Before the availability of marker data, breed proportion estimates could only be obtained from pedigree information. Single nucleotide polymorphism (SNP) genotypes potentially allow for more accurate estimates of breed proportion, even in the absence of pedigree records. A number of tools exist for predicting breed composition using genetic markers. Most of these implement statistical methods developed for prediction of admixture levels and use the complete set of markers. Common approaches are based on hidden Markov Model (HMM) clustering algorithms or maximum likelihood procedures (Frkonja *et al.*, 2011). To obtain estimates of breed composition in crossbred populations, a *reference population* consisting of genotypes from purebred animals that may have contributed to the composite population are used. Dodds *et al.* (2014) explored genomic selection methodology by comparing GBLUP with regression methods to develop predictions for breed proportions. This study showed that either method can be applied

but which one is better depends on the structure of the ancestral breeds that contributed to the population of interest. Blackburn *et al.* (2014) showed that, in composite populations, using a small set of 60K markers (extracted from the Bovine HD SNP chip) at high frequency in each of the founder breeds; the proportion of the founder breeds in the composite animals can be estimated. In combination these studies showed how promising SNP panels are to characterize genetic composition within a population; nevertheless if the main objective is product certification and breed verification, the use of a full high density SNP panel has economic constraints. Consequently, it is of practical importance to find a small and accurate subset of SNP to estimate breed composition. In the present study we explored the use of Differential Evolution to identify a small SNP panel that can accurately be used for Hanwoo breed composition evaluation.

MATERIAL AND METHODS

Data. Genotype information from the BovineHD (700K Illumina BeadChip) array was available for a total of 2,453 animals from different cattle breeds (Hanwoo, Angus, Brahman, Charolais, Holstein and Jersey). The data set was divided into a discovery (2,253) and a validation (200) population. The discovery and validation samples were mutually exclusive. First, 200 samples were randomly selected among the 6 different breeds previously mentioned as validation samples and then the remaining samples were used as discovery population. After quality control 497,737 SNP across all populations were kept for further analysis. A second dataset consisting of genotype information from 24 Yeonbyun samples was also used to validate the proposed method. Yeonbyun are genetically highly related to Hanwoo (populations separated during the Korean War) with some level of crossing with European breeds (Gondro *et al.*, 2012a); which makes them suitable as a proxy for crossbred Hanwoo.

Breed proportion. Breed proportion estimates were calculated using the supervised option with K=7 implemented in the ADMIXTURE software (Alexander *et al.*, 2010). From the breed proportion output we estimated the Hanwoo proportion of the validation set animals. Breed proportion was considered as the *trait*. Phenotypes of pure bred Hanwoo animals were coded as 1; animals of the other reference breeds were coded as 0; therefore prediction of the validation animals using the SNP subset was expected to be in the range of 0–1. A principal component (PC) analysis was also performed to better understand breed composition, to explore potential sub-structure within the sample and for graphic display of the data.

Evolutionary Algorithm. An algorithm based on Differential Evolution (DE) (Storn and Price, 1997) was used to select the best set of SNP for breed proportion estimation. Random keys were used to select the SNP panel. A random key is an evolvable vector of real numbers (one for each SNP) that are sorted in the objective function and then the ranking of the key is used to rank the SNP. The idea is that, SNP for better breed proportion estimation will evolve to higher values in the key and the rest to lower values; once the keys are sorted they reflect the relative value of a given SNP. Predefined *cutoff values* (100, 200, 300, 400, 500, 1000 and 5000) were used to select the number of SNP in the panel. Basically the DE evolves and sorts the SNP based on their key values and uses the top ranked ones up to the number defined by the *cutoff* parameter. More in-depth details on the algorithm are given in Gondro and Kwan (2012b). An objective function was used to find the *fitness* of the selected SNP panel. In the objective function, the discovery population was further divided into two subsets: i) a subset population (1,253 random samples) with known Hanwoo proportion and ii) another subset population (1,000) with unknown Hanwoo proportion (proportions were set to missing for these samples). A genomic relationship matrix (GRM) was calculated using only the selected SNP panel with the all 2,253 discovery samples. The resulting GRM was used to predict the Hanwoo proportion (using GBLUP if number of SNP > number of animals, SNP-BLUP otherwise) for the 1,000 samples with unknown breed estimates. The fitness of a selected SNP panel (accuracy) was defined as the correlation between the

observed and the predicted Hanwoo proportion for these 1,000 samples. The DE evolved for 100 to 500 generations depending on the number of SNP used in the panel; SNP panel size being inversely proportional to the number of iterations. We used 500 generations to evolve the DE for SNP panels with 100 – 400 SNP; 200 and 100 generations for SNP panels with 500 – 1,000 and 5,000 SNP respectively. Once the DE finished, the SNP panel with the highest fitness value was selected and the SNP effects were saved to perform prediction on the validation data. Prediction for Hanwoo proportion was calculated using the following equation: $\hat{y} = 1_n\mu + \sum_i \boldsymbol{X}q_i + e$, where μ is the mean, $\boldsymbol{X}$ is an incidence matrix linking observations to SNP genotypes, q_i is the estimated effect of each SNP and i = 1 to the number of SNP on the SNP panel.

Random subsets. To compare the performance of the DE Algorithm SNP were randomly selected for different panel sizes (100, 200, 300, 400, 500, 1000 and 5000) and then SNP-BLUP was performed on both validation sets. The accuracy of Hanwoo proportion estimates shown for each SNP panel is the average of 10 independent random samples.

RESULTS AND DISCUSSION

Figure 1 shows the first two PC of the genomic relationship matrix applied to 164 animals from different cattle breeds. Hanwoo cattle is clearly separated from the European breeds and Yeonbyun animals are between Hanwoo and European breeds, showing that most of the animals are genetically highly related with Hanwoo cattle, agreeing with Gondro *et al.* (2012a) and that these cattle have potentially been crossed with European breeds. Consequently Hanwoo proportions in Yeonbyun animals are expected to be between 0 and 1. These results were confirmed when calculating the Hanwoo proportion of the validation set using the SNP panel selected with the DE Algorithm (data not shown).

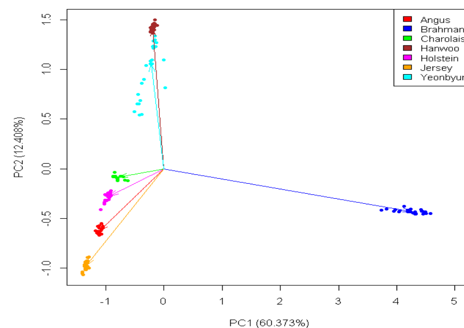


Figure 1. Top 2 axes of variation from principal component analysis of the breeds used to select the marker panel for breed proportion estimates.

Accuracy of breed proportion estimates using different number of markers selected with the DE Algorithm ranged between 0.83 and 0.99 for sets of 100-5000 SNP (Table 1). When using Hanwoo and other European breeds as a validation set, the accuracies didn't change much among SNP subsets using the DE Algorithm (100 to 300 SNP 98% and >300 99%) or random selection (93% with 100 SNP and 96 to 99% with >100 SNP). Results show that the number of SNP included in the different panels is sufficient to extract information about breed proportion in the population, being better than what previous studies suggest (5K SNP, Frkonja *et al.*, 2011; and 60K, Blackburn *et al.*, 2014) and demonstrating that using only a fraction of SNP from the HD SNP panel we could predict the phenotype or Hanwoo proportion which is comparable with the prediction accuracies achieved when all SNP are used (0.99). However if the accuracy is important then we need to use larger SNP panels (i.e. panels with about 1,000 SNP). On the contrary if the cost is the main concern then we could use panels with a lower number of SNP by accepting a small decrease in accuracy. It should be noted though that Hanwoo is genetically quite distinct from European breeds and panels to resolve breed composition within European breeds will probably need to be larger.

Table 1. Accuracy of breed proportion estimates using Differential Evolution (DE) Algorithm and random SNP with different number of markers in the Yeonbyun validation set.

SNP	DE	Random
100	0.83	0.51
200	0.91	0.72
300	0.91	0.76
400	0.91	0.83
500	0.94	0.81
1000	0.94	0.91
5000	0.99	0.98

Knowledge of animal breed composition in livestock populations is also important to identify the best candidates for selection. In crossbred populations it allows effective exploitation of heterosis effects by enabling accurate decisions about the best mattings to be performed within the population. Further, breed composition of crossbred animals in livestock populations provides information on the type and level of crossbreeding as well as on the level of recombination loss (e.g. VanRaden and Sanders 2003). Use of SNP panels increases the level of resolution at which the genetic diversity of composite breeds can be managed. Breed prediction also becomes possible in the case of incomplete or missing pedigrees and in the search for the best type of cross or composite of breeds.

CONCLUSION

The method presented in this study suggests that small, accurate and cost effective SNP panels can be identified for breed proportion evaluation. The results represent a promising approach for product certification and to ensure breed purity in Hanwoo at a low cost. This method can be ported seamlessly to other breeds as well.

AKNOWLEDMENTS

This work was supported by a grant from the Next-Generation BioGreen 21 Program (Project No. PJ01134906), Rural Development Administration, Republic of Korea and an Australian Research Council Discovery Project DP130100542.

REFERENCES

Alexander D.H., Novembre J. and Lange K. (2010) ADMIXTURE 1.04 Software manual, Version 1.04.

Blackburn H.D., Paiva S.R., Sollero B.P., Biegelmeyer P., Caetano A.R. and Cardoso F. (2014) Proc 10th WCGALP. Manuscript 470.

Bovine HapMap Consortium. (2009) *Science.* **324**:528.

Dodds K.G., Auvray B., Newman S.N. and McEwan J.C. (2014) *BMC Gen.* **15**:92.

Frkonja A., Gredler B., Schnyder U., Curik I. and Solkner J. (2012) *Anim. Gen*. **43**:696.

Gondro C., Jang G.W., Lee S.H., Yeon S.H. and Seong H.H. (2012a) Proc. 15th AAAP. Paper Code: C15-OP-161.

Gondro G. and Kwan P. (2012b) in Multidisciplinary Computational Intelligence Techniques: Applications in Business, Engineering and Medicine, pp. 351-377 Editor IGI Global.

Han S.W. and Lee B.O. (2010) *J. Agric. Life Sci.* **22**:73.

Jo C., Cho S.H., Chang J., Nam K.C. (2012) *Anim. Front.* **2**(4):32.

Kim K. H., Kim Y. S., Lee Y .K. and Baik M. G. (2000) *Meat Sci.* **55**:47.

Lee S.H., Park B.H., Sharma A., Dang C.G., Lee S.S. *et al.* (2014) *J. Anim. Sci. and Tech.* **56**:2.

R Development Core Team. (2012) R Foundation for Statistical Computing.

Storn R. and Price K. (1997) *J. Global Optim.* **11**(4):341.

VanRaden P.M. and Sanders A.H. (2003) *J. Dairy Sci.* **86**:1036.

Proc. Assoc. Advmt. Breed. Genet. 21: 477-480

ALLELE SPECIFIC EXPRESSION IS PERVASIVE IN CATTLE

C.J. Vander Jagt[1,4*], A.J. Chamberlain[1,4*], B.J. Hayes[1,2,4], L.C. Marett[1], M.E. Goddard[1,3,4]

[1]Department of Economic Development, Jobs, Transport and Resources, Victoria, Australia
[2]Biosciences Research Centre, La Trobe University, Victoria, Australia
[3]Faculty of Veterinary & Agricultural Science, University of Melbourne, Victoria, Australia
[4]Dairy Futures CRC, Victoria, Australia
[*]These authors contributed equally to this work

SUMMARY

Gene expression can be regarded as a complex trait phenotype, affected by a number of mechanisms, including *cis*-regulatory genetic variation. Allele specific expression (ASE) analysis can be used to determine the importance of *cis*-regulatory variation. In this study, using RNAseq data mapped to parental reference genomes, we analyse the ASE patterns of 17 tissue types and white blood cells (WBC) taken from a single lactating dairy cow. We found that 76% of all heterozygous single nucleotide polymorphisms (SNPs) tested (total 25,251) had significant ($p<0.01$) ASE in at least one tissue type and of all tested genes containing more than 1 tested SNP (7,985), 74% contained greater than 1 ASE SNP. However, there is a large variation between tissues in which genes contain SNP displaying ASE. We conclude that ASE is pervasive in cattle. Identification of these ASE SNP will aid in the detection of *cis*-regulatory variants responsible for phenotypic variation in bovine production traits, which in turn, may lead to improved selection of animals.

INTRODUCTION

Detection of ASE depends on the ability to differentiate the gene product of one parental chromosome from that of the other, and then to quantitate the relative amounts of each gene product. Using RNAseq data, this can be achieved by examining the imbalance of parental alleles expressed at heterozygous SNP (Pastinen 2010). When only one parental allele is expressed at a known heterozygous SNP, it may be indicative of gene imprinting. ASE complements the more traditional expression quantitative trait loci (eQTL) data, narrowing genomic regions of interest and has been successful in helping pin-point causative variants (Ge *et al.* 2009; Montgomery *et al.* 2010; Pickrell *et al.* 2010). The variants used to measure ASE from RNAseq data are within transcribed regions, nevertheless, identification of those ASE SNP in mRNA may serve as markers for the existence of causal regulatory variants close by. It is the identification of these causal regulatory variants affecting quantitative traits in livestock species that are of most interest, as a subset of these mutations could affect traits in the breeding goals for these species.

In this paper we present the results of an allele specific expression analysis of 17 tissues and WBC taken from a lactating Australian Holstein cow at a single point in time. This cow and her sire were sequenced as part of the 1000 bull genomes project and thus phased genotypes of all her heterozygous variants were available to create parental genomes. Alignment to parental genomes is considered the most accurate mapping method and least likely to result in mapping bias (Degner *et al.* 2009). Results of this study indicate pervasive ASE in bovine and large variation between tissues in which genes display ASE.

METHODS

100 base paired end RNA-seq reads were generated on an Illumina HiSeq2000 from 17 different tissues and WBC (in triplicate - see Table 1, column 1 for tissue types) taken from a single lactating Australian Holstein cow (25 months old, 65 days into first lactation). Reads per

tissue ranged from 40 to 100 million. Maternal and paternal reference genomes were created by editing UMD3.1 bovine genome assembly at all heterozygous variant sites from this cow using phased genotypes from 1000 bull genomes run 3 (Daetwyler *et al.* 2014). Paired RNA reads for each tissue replicate were aligned twice, once to each parental reference genome, using TopHat2 (Kim *et al.* 2013) and Ensembl release 75 genome annotation, allowing for two mismatches. Alignment files for each tissue replicate were merged, sorted and indexed using SAMtools (Li *et al.* 2009). Maternal and paternal allele counts for known heterozygous SNP for this cow (Daetwyler *et al.* 2014) were extracted using SAMtools mpileup (version 0.1.14). SNP were then filtered to only consider those falling within gene exon boundaries, with a minimum read depth of 10 in both parental reference alignments and the most abundant allele in both the maternal and paternal alignments had to agree (removing SNP falling in regions with obvious mapping bias). SNP were considered as having significant (p<0.01) ASE using the following Chi-squared (χ^2) test:

$$\chi^2 = \frac{\left(\frac{\left(r_m a_p - a_m r_p\right)^2 N}{ramp}\right)}{2}$$

where r was the count of reference alleles aligned to both parental genomes, a was the count of alternate alleles aligned to both parental genomes, m was the count of reference and alternate alleles aligned to the maternal genome, p was the count of reference and alternate alleles aligned to the paternal genome, r_m was the count of reference alleles aligned to the maternal genome, r_p was the count of reference alleles aligned to the paternal genome, a_m was the count of alternate alleles aligned to the maternal genome, a_p was the count of alternate alleles aligned to the paternal genome and N was the total number of alleles aligned to both parental genomes. Chi-squared values were divided by 2 to account for the value of N being derived from the counts of both parental haplotypes.

RESULTS AND DISCUSSION

Figure 1 demonstrates that there is little bias toward the reference allele for all SNP tested, with reference allele frequency normally distributed and centred at 0.5, indicating that our strategy of mapping reads separately to parental genomes was largely successful. As will be discussed later, lung has a large number of ASE SNP. Figure 1 also reveals a large number of SNP have extreme ASE (peaks at 0 and 1), however there is some bias in the SNP that display a reference allele frequency of 1. We believe this is due to errors in the whole genome sequencing of this cow, resulting in SNP called heterozygous when in fact the cow is homozygous at that position.

In total 25,251 SNP were tested for ASE in at least one tissue, and these SNP fell within 7,985 annotated genes. 89% of genes tested had significant ASE in at least one tissue (Table 1). Wang et al. (2014) state genes that have multiple SNP supporting ASE have a higher rate of successful verification. Therefore, we also tested the proportion of genes with >1 SNP with significant ASE where the gene had >1 SNP tested, this was 74% (Table 1). These results suggest that between 74-89% of genes show ASE in at least one tissue. This estimate is higher than the majority of published mouse and human literature of 4-53% (Yan *et al.* 2002; Bray *et al.* 2003; Pant *et al.* 2006; Serre *et al.* 2008; Vidal *et al.* 2011; Gao *et al.* 2012; MacEachern *et al.* 2012), though it must be acknowledged that these estimates are for single or few tissues or cell lines.

For individual tissues, the proportion of genes showing significant ASE varied from as low as 8-16% of those tested in thymus, to as high as 71-82% tested in lung. Pant et al. (2006) previously reported that 53% of genes tested showed significant ASE in a study limited to testing only 1,389 genes in WBC, our estimate of 21-33% of genes tested in WBC was lower, however we tested

more genes (4,680). Gao et al (Gao *et al.* 2012) reported 30% of the 8,779 genes tested in human mammary epithelial cells lines showing significant ASE, this also corresponds well to 16-31% of the 3,566 genes tested in our study showing significant ASE in mammary gland. The result of 71-82% of genes tested showing significant ASE in lung seems high, however to our knowledge this is the first time ASE has been tested in lung. Our estimate of 14-25% and 14-26% of the 5,462 and 5,946 genes tested in brain caudal lobe and brain cerebellum respectively are much lower than the estimate of 89% by Crowley et al (Crowley *et al.* 2015) in whole mouse brain, however they had an extremely powerful design testing greater than 12,000 genes in 96 mice from all possible pairwise crosses between the three divergent inbred lines. The power of our study comes from testing many tissues. Interestingly, a recent study undertaken by the GTEx Consortium (2015), compared between-sample and between-tissue sharing of ASE in humans with overall similarity of gene expression. They found that gene expression levels were determined by tissue, and that individuals clustered by tissue. However, allelic ratios have a higher correlation among tissues from the same individuals than among individuals for the same tissue. This suggests that ASE is primarily determined by the individual's genome. Therefore we have likely underestimated the total number of genes displaying ASE in the cattle population, and that further testing in more individuals (currently underway) will uncover more genes that undergo cis-regulation.

This study demonstrates that ASE is pervasive in cattle, supporting the argument by Pai et al. (2015) that much of the variation seen in quantitative traits is likely due to these changes in expression, and that these genes are under *cis*-regulatory control. Attention must now turn to the identification of these cis-regulatory variants. The identification of causative regulatory variants could then be used in livestock genomic selection programs leading to more accurate genomic breeding values and increases in the rate of genetic gain for economically important traits.

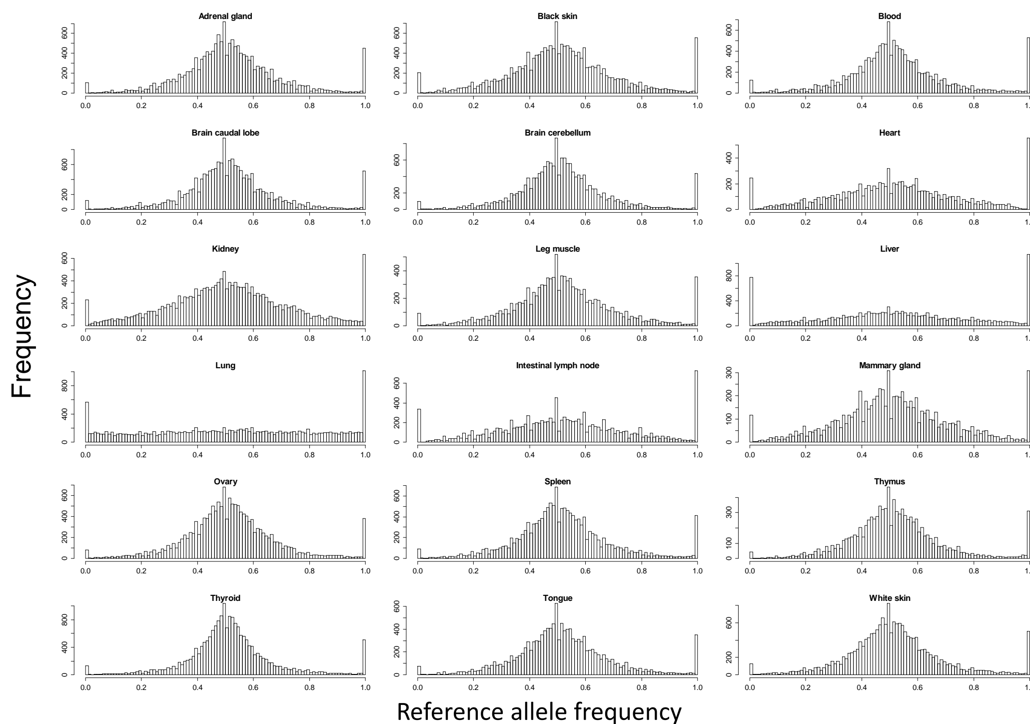


Figure 1. Reference allele frequency distributions for each tissue and WBC.

Table 1. Allele specific expression analysis results

Tissue	# SNP tested	# ASE SNP (% tested)	# Genes tested	# Genes w/ >1 SNP tested	# Genes w/ ASE SNP (% tested)	# Genes w/ >1 ASE SNP (% tested)
Adrenal	14,698	2,636 (18%)	5,462	3134	1,635 (30%)	536 (17%)
Brain caudal lobe	16,594	2,419 (15%)	5,946	3483	1,478 (25%)	494 (14%)
Brain cerebellum	15,460	2,324 (15%)	5,650	3269	1,470 (26%)	466 (14%)
Heart	9,545	2,919 (31%)	3,999	2118	1,869 (47%)	618 (29%)
Intestinal lymph	11,719	3,554 (30%)	4,684	2542	2,391 (51%)	782 (31%)
Kidney	16,616	7,442 (45%)	5,925	3457	3,958 (67%)	1,751 (51%)
Leg Muscle	11,401	2,006 (18%)	4,455	2467	1,394 (31%)	402 (16%)
Liver	12,507	6,773 (54%)	4,887	2715	3,574 (73%)	1,612 (59%)
Lung	14,238	9,216 (65%)	5,419	3032	4,448 (82%)	2,157 (71%)
Mammary	8,161	1,543 (19%)	3,566	1838	1,100 (31%)	302 (16%)
Ovary	15,108	2,043 (14%)	5,588	3229	1,407 (25%)	399 (12%)
Skin black	16,255	4,507 (28%)	5,870	3386	2,776 (47%)	999 (30%)
Skin white	17,087	3,533 (21%)	6,004	3531	2,156 (36%)	766 (22%)
Spleen	14,495	2,066 (14%)	5,317	3071	1,448 (27%)	382 (12%)
Thymus	9,781	986 (10%)	3,981	2159	634 (16%)	182 (8%)
Thyroid	18,181	3,279 (18%)	6,196	3703	2,013 (32%)	688 (19%)
Tongue	12,744	1,671 (13%)	4,850	2718	1,177 (24%)	327 (12%)
White blood cells	12,768	2,662 (21%)	4,680	2690	1,543 (33%)	552 (21%)
Total	25,251	19,082 (76%)	7,985	4856	7,067 (89%)	3,570 (74%)

REFERENCES

Bray N.J., Buckland P.R., Owen M.J. & O'Donovan M.C. (2003) *Human Genetics* **113**: 149-53.
Crowley J.J., Zhabotynsky V., Sun W, *et al.* (2015) *Nature Genetics* Advanced online article.
Daetwyler H.D., Capitan A., Pausch H., *et al.* (2014) *Nat Genet* **46**: 858-65.
Degner J.F., Marioni J.C., Pai A.A., *et al.* (2009) *Bioinformatics* **25**: 3207-12.
Gao C., Devarajan K., Zhou Y., *et al.* (2012) *BMC Genomics* **13**: 570.
Ge B., Pokholok D.K., Kwan T., *et al.* (2009) *Nature Genetics* **41**: 1216-22.
GTEx Consortium (2015) *Science* **348**: 648-60.
Kim D., Pertea G., Trapnell C., *et al.* (2013) *Genome Biology* **14**: R36.
Li H., Handsaker B., Wysoker A., *et al.* (2009) *Bioinformatics* **25**: 2078.
MacEachern S., Muir W.M., Crosby S.D. & Cheng H.H. (2012) *Frontiers in Genetics* 2.
Montgomery S.B., Sammeth M., Gutierrez-Arcelus M., *et al.* (2010) *Nature* **464**: 773-U151.
Pai A.A., Pritchard J.K. & Gilad Y. (2015) *PLoS Genetics* **11**: e1004857.
Pant P.V.K., Tao H., Beilharz E.J., *et al.* (2006) *Genome Research* **16**: 331-9.
Pastinen T. (2010) *Nature Reviews Genetics* **11**: 533–8
Pickrell J.K., Marioni J.C., Pai A.A., *et al.* (2010) *Nature* **464**: 768.
Serre D., Gurd S., Ge B., *et al.* (2008) *Plos Genetics* **4**.
Vidal D.O., De Souza J.E.S., Pires L.C., *et al.* (2011) *Genome* **54**: 120.
Wang X. & Clark A.G. (2014) *Heredity* **113**: 156.
Yan H., Yuan W., Velculescu V.E., Vogelstein B. & Kinzler K.W. (2002) *Science* **297**: 114.

Proc. Assoc. Advmt. Breed. Genet. 21: 481-484

A MAP OF BOVINE LONG NON CODING RNA ACROSS 18 TISSUES

Lambros T. Koufariotis[1,2,3], Yi-Ping Phoebe Chen[1], Amanda Chamberlian[2], Christy Vander Jagt[2] and Ben J. Hayes[1,2,3]

[1]College of Science, La Trobe University, Melbourne, VIC 3086, Australia
[2]Department of Environment and Primary Industries, AgriBio, Bundoora, VIC 3086, Australia
[3]Dairy Futures Co-operative Research Centre, 5 Ring Road, Bundoora, VIC 3086, Australia

SUMMARY

Long non-coding RNA (lncRNA) are common elements in vertebrates and other lesser organisms that possess numerous regulatory and cellular roles. Long ncRNA are well characterized in humans and mice, however in other species, there is comparatively little information of these elements. Identifying lncRNA in bovine could aid in identifying additional sites in the genome where mutations are likely to contribute to variation in complex traits along with understanding the evolutionary importance and constraints of these transcripts. This is important in bovine, since genomic predictions are increasingly used for genetic improvement of milk and meat production. We address the main challenge in identifying lncRNA, namely distinguishing lncRNA transcripts from unannotated genes, by developing a strict lncRNA filtering pipeline. Our aim was to identify and annotate novel lncRNA transcripts in the bovine genome captured from RNA Sequencing (RNA-Seq) data across 18 tissues, sampled in triplicate. We find 9,886 transcripts passed strict filtering criteria and show moderate to high expression. Further we find many unique lncRNA transcripts are downregulated in a tissue specific manner. This study also identified a large number of novel unknown transcripts in the bovine genome, many having high protein coding potential, indicating a clear need for better annotations of protein coding genes in the bovine genome.

INTRODUCTION

The mammalian genome is highly complex, with protein coding genes considered some of the most important elements within the genome, however these only account for only a small portion of the entire transcriptome. It has recently been revealed that about 1-2% of the human genome is transcribed to messenger RNA (mRNA) (Frith *et al*. 2005) and up to 50% of the transcribed genome does not align to known protein coding regions (Hung and Chang. 2010). It is hypothesized that these non-protein coding RNA can either be transcriptional artifacts due to RNA Polymerase II errors in elongation (Van Bakel *et al*. 2010) or non-coding RNA (Kapranov *et al*. 2010). Evidence is accumulating for the later hypothesis, with studies across a range of species, including humans (Cabili *et al*. 2011), mouse (Dinger *et al*. 2008) and bovine (Qu and Adelson. 2012, Weikard *et al*. 2013) finding many novel ncRNA across a range of tissues.

Recent advances in transcriptome sequencing has allowed for the discovery of a new class of non-coding RNA transcripts that are surprisingly long, known as long noncoding RNA (lncRNA) (Marques and Ponting. 2014). Long noncoding RNA are classified as having an arbitrarily defined length of more than 200 nucleotides with weak or no protein coding potential and generally have lower expression levels than mRNA (Marques and Ponting. 2014). Functions of lncRNA are quite diverse, but some of the better studied lncRNA have described functions in regulating and guiding epigenetic marks and gene expression. These elements are coded almost anywhere in the genome including intergenic regions (also known as long intergenic ncRNA (Qu and Adelson. 2012). One of the best studied examples is *Xist,* a gene responsible for facilitation of imprinting the X chromosome that is in fact a lncRNA (Clemson *et al*. 1996).

While there have been a few studies in bovine isolating novel lncRNA (Weikard *et al*. 2013,

Billerey *et al*. 2014) there is still comparatively little information when compared to the repertoire of lncRNA found in human and mouse genomes. In this study we describe a comprehensive catalogue of putative bovine lncRNA expressed in 18 tissues and located within intergenic regions. Given the main challenge in identifying lncRNA is distinguishing them against transcripts from unannotated genes, we used stringent filtering methods to discriminate potentially protein coding RNA from ncRNA, acknowledging that the stringent filters may discard some true lncRNAs. We also compared our putative lncRNA to catalogues from mouse and human, to gain insights into the evolution of lncRNA across species. This information is of particular value since mutations that might be found within these lncRNA elements can potentially contribute to variations in complex traits.

MATERIALS AND METHODS

RNA extraction, tissue sampling, sequencing and alignment. The tissues used in this study include: adrenal gland, black skin, white blood cells, caudal lobe of brain, brain cerebellum, heart, kidney, leg muscle (semimembranosus), liver, lung, intestinal lymph node, mammary gland, ovary, spleen, thymus, thyroid, tongue and white skin.

The quality control, filtration, read alignment to the reference genome and generation of the SAM files for the 18 tissue samples were performed as described in another study (Chamberlain *et al*. 2014).

Finding intergenic long noncoding RNA. We used a Cufflinks/Cuffmerge/Cuffcompare pipeline to assemble transcripts for all three technical replicates in each tissue sample to the Ensemble reference gene set release 75. Entries that had a class code of either “u”, (unknown intergenic transcript), or “x”, (exonic overlap with the reference genome but on the opposite strand) were extracted and kept for further analysis. Similar to (Weikard *et al*. 2013) we used Cuffcompare to compare our transcripts to those in the NCBI iGenomes repository to filter out transcript with protein sequences, giving us a total of 47,117 transcripts with unknown annotations. We used the UCSC utility twoBitTofa to obtain the nucleotide sequences for the transcripts.

Long non-coding RNA filtering pipeline. To find transcripts most likely to be noncoding RNA transcripts, we developed a 3 stage pipeline to filter out the transcripts that had a high chance of having protein coding potential.

Stage 1. ORF Analysis. getorf from the EMBOSS software package was used to find all possible open reading frames (ORF) in all directions of the transcript. We performed a blastp search on all ORF sequences to determined possible protein coding domains using an E-value of 1e-06 as cut-off. If no significant sequence matches were determined then the transcript was considered to be a potential lncRNA.

Stage 2. Blastx. We determined if our transcripts had any significant matches with protein sequences by using the tool blastx. An E-value of 1e-06 was used as cut-off. Only transcripts that did not show any significant matches with known protein coding sequences were considered.

Stage 3. CPC Tool. The third stage used the tool Coding Potential Calculator (CPC) which predicts the coding and noncoding potential of a transcript. We selected for transcripts as potentially noncoding if they have a score of < -0.5.

Read counts, filtering of low read counts and differential expression analysis. Read counts were obtained using the tool HTSeq and was run with default parameters only specifying for non-stranded (--stranded=no) and union mode (--mode=union) to get the counts matrix for each unkown transcript across all tissues and replicates. The final counts matrix file was used as input for the tool EdgeR for normalization and for filtering transcripts that had very low read counts (read count <25 across all three replicates for each tissue).

Differential analysis was carried out by performing a t-test for each tissue with all the other

tissue samples. The standard error was calculated by subtracting the mean across all tissue with the mean for each tissue. If the t-test had a *P*-value of <0.05 and a positive standard error, then the transcript was considered to be upregulated. If the t-test had a *P*-value <0.05 and a negative standard error then that transcript was considered to be downregulated. All other transcripts were considered to have no differential expression.

Homology analysis with ncRNA in human and mouse. Human and mouse ncRNA were obtained from; GENCODE v7, NONCODE v4 and lncRNAdb databases. A blastn search was performed using an E-value of 1e-06 to blast the unknown transcripts with the human or mouse databases. From this we extracted the transcripts that had significant matches with a known lncRNA.

RESULTS AND DISCUSSION

After transcript assembly and annotation of RNA-Seq reads a total of 47,117 transcripts that aligned to the genome but did not align to protein coding genes or had protein coding annotations were found. These assembled transcripts were passed through the filtering pipeline to determine coding or noncoding potential. We defined putative lncRNA only if the transcripts passed all 3 stages of the filtering pipeline (methods) and had moderate to high expression levels after filtering for low read counts with EdgeR. A total of 9,886 putative lncRNA passed all three filters and were considered for further analysis.

We find that tissues involved in similar organ functions share very similar expression of putative lncRNA. These correlations are lower that what we find in the protein coding genes from the same datasets (Chamberlain *et al*. 2014). The expression patterns of our putative lncRNA show that 37% are downregulated, while 4% are upregulated and 59% show no differential expression (Figure 1).

The vast majority of the lncRNA are found to be within intergenic regions of the genome, however we do find a total of 1,501 lncRNA (about 15% of total lncRNA) that are located either near the 5’ or 3’ end of protein coding genes or are located within 5 kilobases upstream or downstream of protein coding genes. Due to the lack of stranded information, it is difficult to attempt to identify independently coded transcripts that are coded in the opposite direction of the neighbouring gene. Therefore we measured the concordance of expression between the lncRNA transcript and the neighbouring protein coding gene. A Pearson’s and Spearman’s rank correlation analysis showed that many lncRNA had high correlations with their neighbouring genes, and therefore could be unannotated exons, however a significant minority show no correlations, these may indicate independently coded transcripts.

Comparative analysis with human and mouse lncRNA. To identify putative lncRNA that show sequence conservation we performed a blastn search between our lncRNA and the lncRNA in both human and mouse lncRNA databases. Of the 9,886 lncRNA, only 289 show significant sequence similarities with known human lncRNA and 119 show significant sequence similarities with known mouse lncRNA. Further, only 36 putative lncRNA show sequence similarities with both a human and mouse lncRNA. Long ncRNA were also compared to other bovine lncRNA found in similar studies using either pigmented or non-pigmented skin cells (Weikard *et al*. 2013) or bovine muscle cells (Billerey *et al*. 2014). Of the catalogue of lncRNA in the skin cells we find 848 (out of 4,948) lncRNA that overlap with our catalogue of lncRNA. Of the 584 lncRNA found in muscle cells, we find a total of 129 that overlap with our lncRNA. Due to the fact that lncRNA are tissue specific and also can be expressed in different developmental stages we acknowledge that these catalogues provide valuable information of potential lncRNA in the bovine genome. Further, studying these regions will assist in finding new classes of genes that, while lacking the ability to code for proteins, can have mutations that could potentially affect complex dairy traits of interest, such as milk volume, fat percent, protein percent and mammary system.

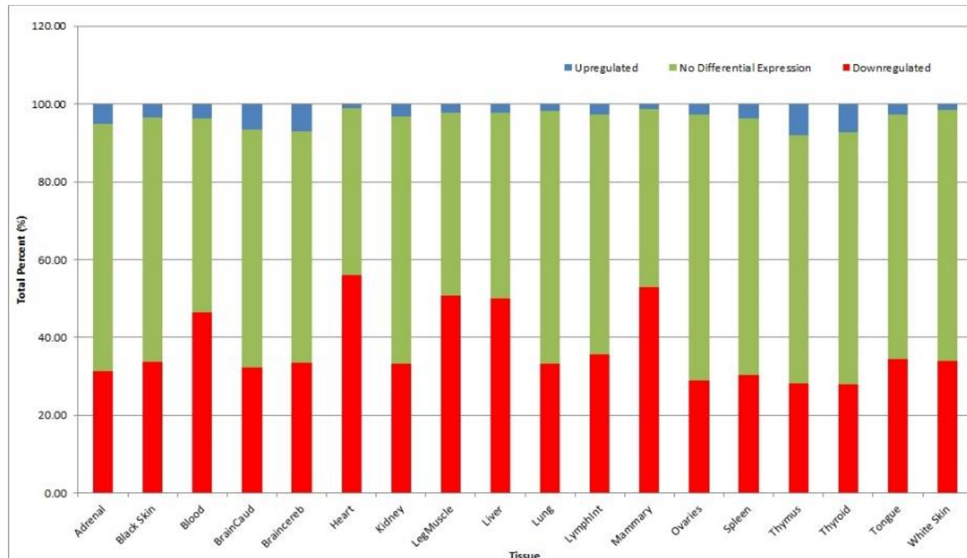


Figure 1. Percent of lncRNA that are upregulated, downregulated or not differentially expressed. Red bars indicate percent of downregulated lncRNA for each tissue. Blue bars indicate percent of upregulated lncRNA for each tissue. Green bars indicate no differential expression.

REFERENCES

Billerey C., Boussaha M., Esquerre D., Rebours E., Djari A., Meersseman C., Klopp C., Gautheret D. and Rocha D. (2014) *BMC genomics* **15**:499.

Cabili M.N., Trapnell C., Goff L., Koziol M., Tazon-Vega B., Regev A. and Rinn J.L. (2011) *Genes & Development* **25(18)**:1915-1927.

Chamberlain A., Jagt C.V., Goddard M. and Hayes B. (2014). A gene expression atlas from bovine RNAseq data. *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*.

Clemson C.M., McNeil J.A., Willard H.F. and Lawrence J.B. (1996) *The Journal of cell biology* **132(3)**:259.

Dinger M.E., Amaral P.P., Mercer T.R., Pang K.C., Bruce S.J., Gardiner B.B., Askarian-Amiri M.E., Ru K., Solda G., Simons C., Sunkin S.M., Crowe M.L., Grimmond S.M., Perkins A.C. and Mattick J.S. (2008) *Genome Research* **18(9)**:1433-1445.

Frith M.C., Pheasant M. and Mattick J.S. (2005) *European journal of human genetics EJHG* **13(8)**:894-897.

Hung T. and Chang H.Y. (2010) *RNA biology* **7(5)**:582-585.

Kapranov P., St Laurent G., Raz T., Ozsolak F., Reynolds C.P., Sorensen P., Reaman G., Milos P., Arceci R., Thompson J. and Triche T. (2010) *BMC Biology* **8(1)**:149.

Marques A.C. and Ponting C.P. (2014) *Current opinion in genetics & development* **27c**:48-53.

Qu Z. and Adelson D.L. (2012) *PloS one* **7(8)**:e42638.

Van Bakel H., Nislow C., Blencowe B.J. and Hughes T.R. (2010) *PLoS biology* **8(5)**:e1000371.

Weikard R., Hadlich F. and Kuehn C. (2013) *BMC genomics* **14**:789.

Proc. Assoc. Advmt. Breed. Genet. 21: 485-488

POTENTIAL ROLE OF lncRNA CYP2C91-PROTEIN INTERACTIONS ON IMMUNE DISEASES AND OBESITY

P. Suravajhala[1], L.J.A. Kogelman[1], G. Mazzoni[1] and H.N. Kadarmideen[1]

[1] Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 7, 1870 Frederiksberg, Denmark

SUMMARY

With unprecedented increase in next generation sequencing (NGS) technologies, there has been a persistent interest on transcript profiles of long noncoding RNAs (lncRNAs) and protein-coding genes forming an interaction network. Apart from protein-protein interaction (PPI), gene interaction models such as Weighted Gene Co-expression Network Analysis are used to functionally annotate lncRNAs in identifying their potential disease associations. To address this, studies have led to characterizing transcript structures and understanding expression profiles mediating regulatory roles. In the current analysis, we show how a lncRNA- cyp2c91 contributes to the transcriptional regulation localized to cytoplasm thereby making refractory environment for transcription. By applying co-expression network methods and pathway analyses on genes related to a disease such as obesity from F2 pig model, we show that we can gain deeper insight in biological processes such as the perturbances in immune system, and get a better understanding of the systems biology of diseases. We believe this study has implications for finding prognostic and diagnostic markers for obesity and immune related diseases.

INTRODUCTION

With unprecedented increase in next generation sequencing (NGS) technologies, there has been a persistent interest on transcript profiles of long noncoding RNAs (lncRNAs) and protein-coding genes forming an interaction network. Apart from protein-protein interaction (PPI), gene interaction models such as Weighted Gene Co-expression Network Analysis (WGCNA; Xue *et al.,* 2013) are used to functionally annotate lncRNAs in identifying their potential disease associations (Cogill and Wang, 2014). To address this, studies have led to characterizing transcript structures and understanding expression profiles mediating regulatory roles and comparing them with the ENCODE project (The ENCODE project. 2014). Recent reports show how lncRNAs contribute towards regulatory interactions with their non-coding peers like miRNAs (Jalali *et al*., 2013). Whether or not lncRNA-protein networks restrain interactions is little known and not detailed. How such regulatory interactions between classes of lncRNAs and proteins would have a significant influence on the organism remains a challenge.

Earlier, we have shown three regulatory genes, viz. CCR1, MSR1 and SPI1 associated with diseases like obesity and osteoporosis using gene network algorithms WGCNA and Lemon-Tree (Kogelman *et al.,* 2014a) applied to NGS-based RNAseq datasets from porcine model for obesity. These clusters of highly co-expressed genes were ranked as highly significant based on their association with obesity-related phenotypes in a F2 pig model (Kogelman *et al* 2014b). With a wide range of biological processes effectively used as regulatory molecules, we anticipate (a) if the coexpressed genes have interacting partners with any long noncoding RNAs (lncRNA), (b) if so, whether or not they affect the coexpression, consequently further changing the networks and influencing the organismal phenotype or disease outcomes, or (c) if not, what would be the outcome of such lncRNA-dependent transcription. From a putative interaction network, we have established functional classes based on several different methods, explicitly focusing on the edge-betweeness, pearson coefficient of overlapping genes, two nearest non-overlapping genes on either side, presence of subcellular location signals (not shown). These resilient methods would distinguish probability of lncRNA to show association/disassociation paradigm, RNA binding

protein-lncRNA interactivity and importantly disease association, if any.

MATERIALS AND METHODS

In the current study, we made a human concordant network from our previous WGCNA result from an animal model (*see Figure 1;* F2 pigs, Kogelman *et al.,* 2014b) and found that 340 of 540 porcine genes have orthologue peers in humans (Figure 2, panel a). The absence of orthologs in human is in agreement with the homology data available from the Pig Analysis Database (PAD) which specifies that about 73% of the sequences are covered by the both genomes (See PAD web reference). From the networks and GenBank annotation, we observed that cyp2c91, a lncRNA interacts with a host of regulatory genes. The betweenness centrality of cyp2c91 with the three regulator genes linked to obesity (CCR1, MSR1 and SPI1) was found to be consistent with the association pattern (Figure 2, panel b). With the hypothesis that lncRNA-protein interactions play an important role in regulating post-transcriptional changes and subsequent localization of the transcript, we used RNA-protein interaction predictor (RPI-pred) to predict whether or not the proteins encoded by these genes and the RNA form interaction pairs (Suresh *et al.* 2015). Considering the fact that these small molecules enter the nucleus without regulation, we asked if any gene products are localized extracellular to nucleus.

RESULTS AND DISCUSSION

We observed that among the three regulator genes, CCR1 was found to be localized in cytoplasm (Figure 2, panel c). Encouraged by the outcome that the three have a plausible role of interaction with cyp2c91, we made a reliable interaction network with the mean disassociation based on the betweeness centrality (Figure 2, panel d). We found that MSR1 and CCR1 are found to be interacting with each other while SPI1 was a lone gene without an interaction pair. Nonetheless, the lncRNA-protein interactions were extended with the CCR1-cyp2c91 association mapped from network genes. The study suggests two ways forward. First, the fold change (log 2 expression) can be attributed to lncRNA-dependent transcription. Second, CCR1-cyp2c91 association is significant when compared to MSR1-cyp2c91 and SPI1-cyp2c91 (indicative of p-values, not shown) where the genes are regulatory in nature forming diseased network. The three regulatory genes are associated with obesity and immune system, possibly linking them to Lupus. This is evident by the fact that several of the genes present in the WGCNA modules of Kogelman et al., 2014 (TNIP1, GPSM3, TFEC, TES, KCP, IRF5, TNPO3, ELF1, ITGAM and TNXB, KLF6, AKR1E2) are related to immune system and systemic lupus erythematosus (SLE). This might allow us to use this network as a model for immune response or obesity.

The genome is lengthily transcribed in eukaryotes and it has been known that many transcripts have larger proportion of noncoding components. Although about 66-73% of the porcine genome (including ESTs, genes etc.) is conserved across humans, a considerable set of genes regulate interactions with lncRNAs. Further, a range of transcribed regions might tend to be regulatory and indicative of enhancing non-functional activity. Moving to a broader spectrum of calling them as junk, we asked for evidences on their regulatory potential based on their association with protein-coding genes. Consistent with the interaction networks from porcine model for obesity, subcellular localization of the products of the three protein-coding genes revealed that two are nuclear while one, CCR1 was found to be in cytoplasm. This is again in agreement with the fact that the subcellular fractions of lncRNA differ significantly from each other, with a majority enriched in the nucleus, cytoplasm and ribosomes. These results show that lncRNA-protein interactions are self-regulating and yet they are dependent on organellar specificity. Our exploratory analysis suggests that CCR1-cyp2c91 contributes to the transcriptional regulation localized to cytoplasm thereby making refractory environment for transcription. By applying co-expression network methods and pathway analyses on genes related to a disease such as obesity and systemic lupus

erythamatosis, we show that we can gain deeper insight in biological processes such as the perturbances in immune system, and get a better understanding of the systems biology of diseases. This stresses the possible need of finding genes linked to lncRNA-protein networks and further use them as potential diagnostic markers in animal and human diseases.

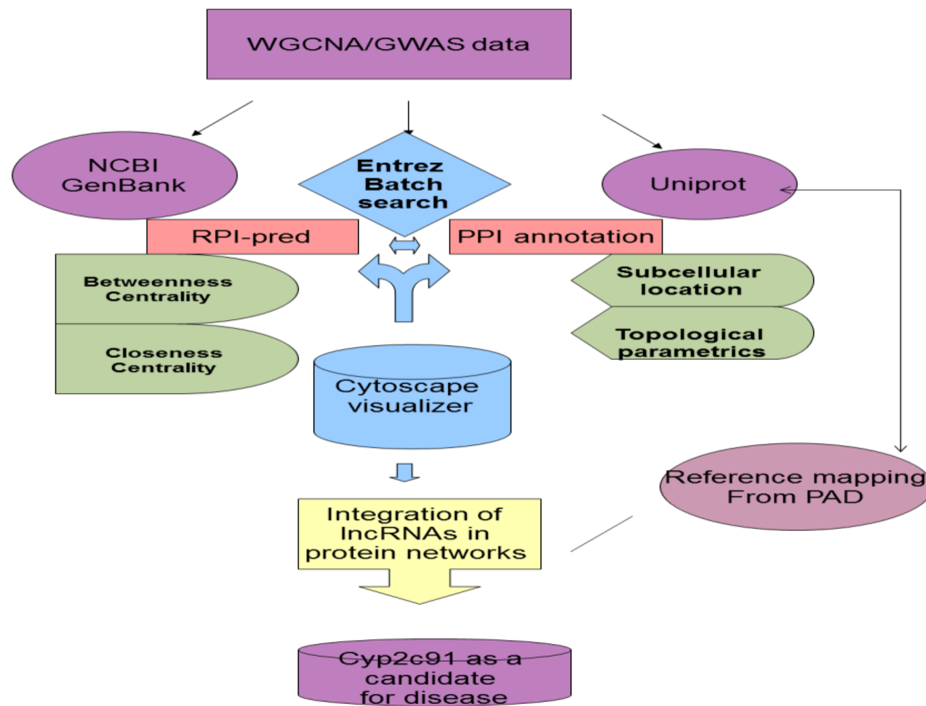


Figure 1: The GWAS data from Kogelman et al. were checked for candidacy across the GenBank. After obtaining the *bona fide* accessions in humans, the sequences were checked using RPI-Pred and protein annotation. The betweenness centrality and closeness centrality values for the nodes were then computed and visualized using Cytoscape. The centrality values are computed for those that do not contain multiple edges. They are the normalized values for each gene/node by dividing the number of pairs of nodes existing in the network. The range would be between 0 to 1 with the condensed values in exponential form as calculated by cytoscape (centrality of vertex). We considered the candidate lncRNA linked to disease after reference mapping and linkage to disease while integrating it into the protein interaction network.

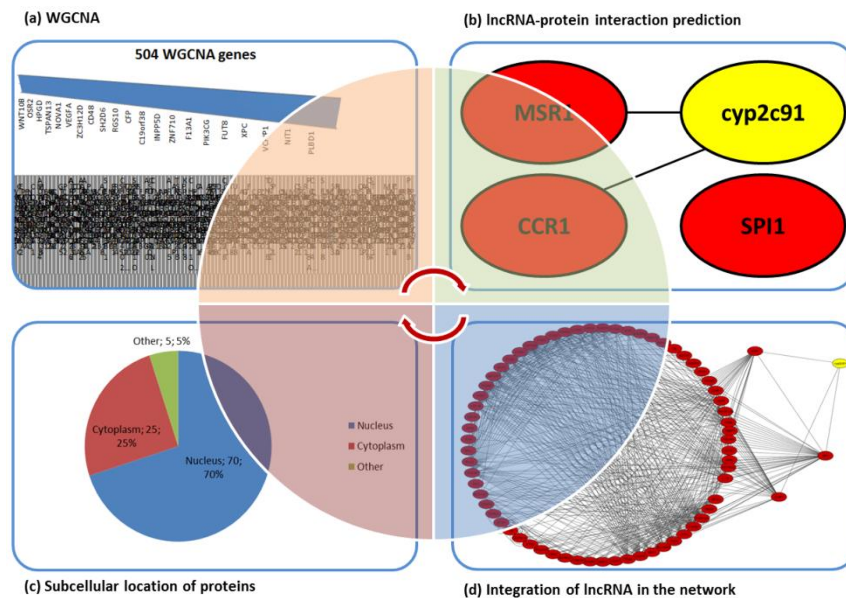


Figure 2: (a) The 504 genes from WGCNA across different modules linked to diseases not limited to obesity and immune response. (b) Representative lncRNA cyp2c91 gene (in yellow) shown to be interacting with three regulatory protein-coding genes (c) Subcellar location of the genes associated in the network and (d) the network topology showing the profiled expression across the regulatory genes associated with cyp2c91. This is indicative of global protein-RNA interaction data.

REFERENCES

Cogill S.B., Wang L. (2014). *Cancer Inform*. ***13***(Suppl 5):49-59.

Jalali S., Bhartiya D., Lalwani M.K., Sivasubbu S., Scaria V. (2013). *PLoS ONE*. ***8***(2): p. e53823.

Kogelman L.J.A., Cirera S., Zhernakova D.V., Fredholm M., Franke L, Kadarmideen H.N. (2014). *BMC Med Genomics*. ***7*:**57.

Kogelman, L.J.A., Pant S.D., Fredholm M., Kadarmideen H.N. 2014. *Frontiers in Genetics (Systems biology),* ***5***:214.

PAD: http://www.piggenome.dk/

Suresh V., Liu L., Adjeroh D., Zhou X. (2015). *Nucleic Acids Res*. ***43***(3):1370-9.

The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, 2004. ***306***(5696): p. 636-40.

van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E et al. (2014). *Genome Biol.* **15**(1):R6.

Xue Z., Huang K., Cai C., Cai L., Jiang C.Y., Feng Y. et al. (2013). *Nature*. ***29***;500(7464):593-7.

Proc. Assoc. Advmt. Breed. Genet. 21: 489-492

CONNECTING GENE EXPRESSION AND PHENOTYPE – PRELIMINARY RESULTS FROM RNA SEQUENCING OF 150 LAMBS

S. Bolormaa[1], R. Behrendt[1], M.I. Knight[1], B.A. Mason[1], C.P. Prowse-Wilkins[1], L. Slocombe[1], C.J. Vander Jagt[1], B.J. Hayes[1,2], A.J. Chamberlain[1] and H.D. Daetwyler[1,2]

[1] Department of Economic Development, Jobs, Transport and Resources, Victoria, Australia
[2] School of Applied Systems Biology, La Trobe University, Victoria, Australia

SUMMARY

Gene expression analysis can aid in prioritising regions or classes of variants for genomic prediction and they increase our understanding of quantitative traits. The number of reads from RNA sequencing that align to a gene can be used to quantify gene expression. We sampled liver and muscle tissues of 150 lambs at slaughter. Their dams had been managed to high, medium, and low body condition scores (BCS) during mid-to-late pregnancy and the lambs were fed three different finishing diets. Differential expression of genes (DEG) was investigated contrasting tissue, BCS, lamb diets, other treatment differences, as well as high and low lamb carcass eye muscle width (CEMW). A large number of DEG were identified between tissues, but only the low versus high BCS comparison resulted in DEG for treatments. DEG were also found when we contrasted high and low CEMW. A strong trend toward down regulation was observed in all tests, except in BCS where all DEG were overexpressed in fatter ewes.

INTRODUCTION

Gene expression refers to the production of RNA transcripts which ultimately result in a gene product. Genes may be expressed only at certain time points in the animal life cycle and expression may be tissue specific. Linking gene expression to measured phenotypes or even to management strategies may help us optimise the production system. Using expression analysis to prioritise candidate genes together with marker or DNA sequence data could ultimately increase genomic prediction accuracies for key traits.

The rate at which genes are expressed can be investigated by measuring the abundance of RNA transcripts in a tissue. Sequencing RNA is an attractive option for this analysis because the number of reads that align to an annotated gene is a reliable measure of abundance (Mortazavi et al. 2008). The analysis of RNA sequence data has a number of uses. At the most basic level, it results in a set of genes expressed in the starting material (i.e. tissue, cell line, etc). Additionally, contrasts can be performed within and across tissues, depending on the variety of tissues sampled or the number of treatments or other phenotypes measured in a study (e.g. liver versus muscle, high versus low body weight). Contrasts across animals may result in sets of differentially expressed genes (DEG), whose degree of difference can be measured via statistical tests and by the Log_2 fold change between classes.

Here we present preliminary results from a differential expression analysis using RNA sequence data on liver and muscle tissues of 150 lambs. Our initial aim was to investigate whether dam body condition score (BCS) during gestation and lamb finishing diets affect gene expression, potentially linking these treatments to genetics.

MATERIALS AND METHODS

The experimental design involved 648 pregnant ewes (Merino x Border Leicester and Maternal/Coopworth Composites) that resulted from artificial insemination to 5 Polled Dorset (PD) and 4 White Suffolk (WS) sires (that had been selected for high, medium, or low lean meat yield genomic breeding values). The core design was based on 3 BCS - ewe live-weight change -

nutritional treatments during gestation that were targeted in ewes reaching BCS2.5, BCS3.0 and BCS3.5 by lambing. The ewe BCS was managed to a flock average of BCS3.0 from prior to artificial insemination and until ultrasound scanning at day 50 to confirm pregnancy. Following scanning, ewes were distributed amongst 18 management cells and nutrition was managed by allocation of feed-on-offer during the last 2 trimesters to achieve the condition score targets. Ewes were maintained within one management group and were given ad libitum access to pasture after lambing to weaning.

Lambs were weaned at 12-13 weeks of age. Male lambs were backgrounded in pasture prior to the finishing with 3 different feedlot diets: 1) high protein, moderate energy, 2) high energy, moderate protein, and 3) high protein and energy. All male progeny (N=436) were slaughtered in three blocks at a commercial abattoir. Full-bone out was performed on 100 lambs, with a partial bone out on the remainder (Pearce et al. 2010). Additionally, a large number or carcass phenotypes were recorded. Here we will present early results on only eye muscle width (CEMW).

Liver and loin muscle tissue samples from slaughter lambs were taken within 10 minutes of death to determine the influence of the experimental treatments on aspects of meat quality, lean meat yield, and gene expression. Samples were flash frozen in liquid nitrogen and stored at -80ºC. 150 lambs (100 full and 50 partial bone outs) randomly selected across all dam lamb nutritional treatments, birth types, breeds and sires, were selected for RNA sequencing. Frozen muscle and liver tissues were ground using the Genogrinder2010 (SPEX). Ground tissue was homogenised in Trizol® (Life Technologies™) and RNA extracted using the Trizol® Plus RNA extraction kit (Life Technologies™). Individually barcoded strand specific RNA sequence libraries were produced using the SureSelect Strand Specific RNA Library Prep Kit (Agilent Technologies). The 300 libraries were combined into one of four pools and 120 bp paired-end sequencing performed on a HiSeq2000 genome analyser (Illumina Inc) with the aim of producing 40 million paired reads per library. Fastq files were called using CASAVA v1.8 (Illumina Inc).

Fastq files were quality controlled using quadtrim (https://bitbucket.org/arobinson/quadtrim) as follows. Low quality bases were trimmed from read ends (phred score < 20). Reads were removed if they: failed the chastity filter, contained more than three Ns, had a mean base quality score of <20, or were < 50% of original read length. Filtered reads were aligned to the SheepOAR3.1 assembly using the Ensembl v78 SheepOAR3.1 (Jiang et al. 2014) annotation file containing 25,202 genes using Tophat2 (Trapnell et al. 2012). A gene by tissue count matrix containing all animal results was generated using the python package HTSeq (Anders et al. 2015). The R program DESeq (Anders and Huber 2010) was used for DEG analysis of tissues, traits and treatments, evaluating whether read counts per gene were significantly different when testing multiple samples belonging to two classes (e.g. liver and muscle) based on a negative binomial regression test. Counts were normalised for mean read depth per sample. The model fitted contrasted two classes across both tissues and within liver or muscle. The following pairwise contrasts were tested: liver and muscle across all samples, liver and muscle in PD and WS sire groups, sire breed within either tissue, 3 BCS levels of dam at gestation, 3 lamb finishing diets, 3 kill days, 3 sire LMY ASBVs, single versus multiple births, and extremes of CEMW (top 10% versus bottom 10%). Genes were reported as DE if their false discovery rate percentage (FDR%) was below 40% at a p value of 0.001 (FDR%=(25202*0.001[number of DEG]$^{-1}$)*100).

RESULTS AND DISCUSSION

An average of 70 million reads (range 19-333 million) were generated for each of 298 samples after filtering for chastity, which is substantially more than the target 40 million. Quality control reduced the reads to 87% of which a mean of 85% of paired reads (range 78-90%) were aligned to the assembly, which was comparable to other sheep studies (Chen *et al*. 2015).

The normalised counts matrix was then used to test contrasts. The comparison of liver and

muscle across all samples yielded 10,116 DEG, which was the largest number in the study (Table 1). It is expected that different genes would be expressed in discrete tissue types. The animals were approximately balanced between PD and WS sire breeds. Analysis of liver versus muscle contrasts within PD and WS revealed 7,990 and 7,060 DEG in each sire breed group, respectively. The DEG overlap between the sire groups was 6,559, indicating that in smaller and independent samples the same genes are expressed. Single breed analysis resulted in negligible additional DEG when compared with the analysis across all animals (data not shown). Contrasts of sire breed resulted in a non-significant small number of DEG. The FDR% of liver versus muscle comparisons was always < 1%. Filtering DEG by increasing Log_2 fold change stringency gradually reduced the number of significant genes. However, a large number resulted in Log_2 fold changes of => 10 (e.g. all samples liver vs muscle, 621). The majority of these DEGs exhibited no expression in one tissue and strong expression in the other. This may be suspicious with small sample size, but in our case of testing 149 lambs it is unlikely that all libraries would miss true transcripts in these genes. DEG in muscle were much more likely to be down regulated (~80-90%) than up regulated, and the percentage of down regulation decreased slightly with increasing Log_2 fold changes before plateauing at ~80% (Figure 1).

Table 1. Number of DEGs at progressive Log_2 Fold Change thresholds and FDR% at Log_2 Fold threshold 1. BCS2.5vs3.5 of ewes refers to low and high body condition during pregnancy.

LIVER vs MUSCLE	**FDR%**	**Absolute Log_2 Fold Change**									
		1	2	3	4	5	6	7	8	9	10+
All Samples	0.3	10116	6553	3751	2350	1633	1205	972	813	683	621
Polled Dorset	0.3	7990	5565	3336	2156	1506	1151	929	784	683	627
White Suffolk	0.4	7060	5820	3443	2186	1528	1155	933	786	690	627
Overlap PD-WS	-	6559									
TREAMENTS	**FDR%**	**Absolute Log_2 Fold Change**									
TEST WITHIN LIVER		1	2	3	4	5	6	7	8	9	10+
BCS2.5vs3.5 of ewes	37.6	67	11	2	1	1	1	1	1	1	1
TRAITS	**FDR%**	**Absolute Log_2 Fold Change**									
BOTH TISSUES		1	2	3	4	5	6	7	8	9	10+
CEMW-top10vsbot10%	4.1	616	481	67	21	8	6	6	6	6	6
LIVER CEMW-top10vsbot10%	7.4	340	241	63	19	6	5	5	5	5	5

The contrasts across treatments detected fewer DEG. In fact, no significant DEGs were found for treatments in muscle. The negative control of kill day also revealed no effect on expression levels, which increases the likelihood that systematic problems during slaughter and related processing were avoided. The only treatment that exhibited DEG was BCS, where the contrast of low (2.5) versus high (3.5) BCS of the ewe resulted in 67 significant genes in liver of the lambs (Table 1). Interestingly, all DEG were up regulated in BCS3.5 when compared to BCS2.5, potentially linking an increase in body condition of ewes during gestation to up regulation of genes in their offspring at slaughter. This is different to all other expression directions observed in this study.

The contrast in extreme lamb CEMW phenotypes resulted in between 616 and 340 DEG across tissues and within liver, respectively (Table 1). The number of DEGs met our FDR% threshold of 40% up to and including Log_2 fold change 3, which was a stronger signal than for ewe body

condition disparity. In the CEMW tests, all DEG were down regulated, which indicates that a reduction in trait phenotype could be shown to be associated with down regulation of gene expression (Figure 1).

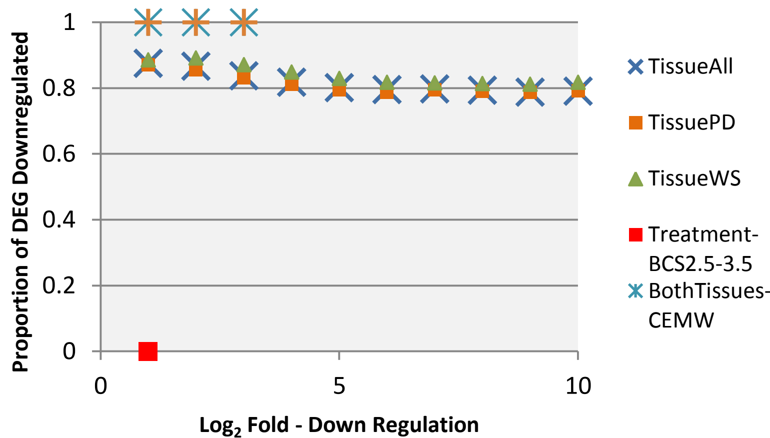


Figure 1. Proportion of down regulated genes at progressing thresholds of Log2 fold change for all significant contrasts, where All is all samples, BCS is body condition score, and CEMW are eye muscle width, respectively. Only points with FDR<40% shown.

KEGG and GO analysis of DEG identified two major gene pathway groups related to fat/cholesterol (liver) and muscle fibre. Several genes (*APOA-1, -2, -4, -5, APOC3*, and *APOF/CAV3*) were involved in cellular cholesterol and phospholipid efflux, homeostasis and transport (Bonferroni $P < 5.5\times10^{-5}$), whereas 5 genes (*ACT, ACTG1, MYL, BMP10*, and *CAV3*) were weakly linked to skeletal muscle differentiation (Bonferroni $P < 8.8\times10^{-1}$). The DEG identified when contrasting dam BCS contained candidate genes involved in stress response and embryonic development (e.g. *KLHDC10, BMP4, MAPKAPK3, ABL1*)

The preliminary analysis of this large RNA sequencing dataset has revealed widespread DE of genes between tissues. It has also connected ewe body condition during gestation to liver gene expression in their lambs. Additionally, differences in CEMW were shown to be mirrored in gene expression patterns. Further analyses on other phenotypic traits and allele specific expression will be performed. The outcomes of this study will contribute towards more precise annotation of the sheep reference genome and it will aid in prioritising genetic markers for genomic prediction.

ACKNOWLEDGMENTS

The authors would like to thank Meat and Livestock Australia for funding the Lamb Foundations Project, and the Growing Food and Fibre Initiative of the Victorian State Government for funding the expression analysis. We are grateful to all staff involved in animal care and the collection of tissues and phenotypes at and after slaughter.

REFERENCES

Anders S, Huber W (2010) *Genome Biology* **11**:R106
Anders S, Pyl PT, Huber W (2015) *Bioinformatics* **31**:166.
Chen HY, Shen H, Jia B, et al (2015) *PLoS ONE* **10**:e0120170
Jiang Y, Xie M, Chen W, et al (2014) *Science* **344**:1168.
Mortazavi A, Williams BA, McCue K et al (2008) *Nat Meth* **5**:621.
Pearce KL, van de Ven R, Mudford C, et al (2010) *Anim. Prod. Sci.* **50**:1107.
Trapnell C, Roberts A, Goff L, et al (2012) *Nat Protocols* **7**:562.

Proc. Assoc. Advmt. Breed. Genet. 21: 493-496

WHAT'S NEXT IN GENOMICS? FUNCTIONAL ANNOTATION OF ANIMAL GENOMES

J. Kijas[1], B. Barendse[1], C. Bottema[2], R. Brauning[3], A. Chamberlain[4], S. Clarke[3], B. Dalrymple[1], R. Tellam[1], B. Hayes[4], J. McEwan[3] and S. Moore[5]

[1] CSIRO Agriculture Flagship, Queensland, Australia
[2] University of Adelaide, South Australia, Australia
[3] Invermay Agricultural Center, AgResearch, Mosgiel, New Zealand
[4] Department of Economic Development, Jobs, Transport and Resources, Victoria, Australia
[5] QAAFI, University of Queensland, Queensland, Australia

SUMMARY

The last five years has witnessed the completion of reference genome projects for each of the major livestock species, along with the application of high throughput SNP genotyping to fast track gene discovery and genomic prediction. This paper explores one possible new direction in genomics and its possible impact on animal science. An international project has been initiated that aims to identify the genomic regions responsible for gene regulation, thereby providing functional annotation of animal genomes (FAANG). This seeks to increase our ability to interpret variation in genome sequence and predict the resulting phenotypic consequence. This has large implications for animal science and in particular animal breeding, given a key objective of genomic prediction is to use molecular data (currently SNP) to predict genetic merit. To successfully annotate the regulatory elements in genomic sequence, the FAANG Consortium has been created to provide coordination and standardisation in data collection, quality control and analysis. Aspects of the consortium are described, along with information on Australia's current and future contributions.

THE GENOME TO PHENOME CHALLENGE

A central goal in biological science involves understanding complex systems, so that accurate predictions about the behaviour of systems can be made. Predictions might involve the susceptibility of an individual to disease or the treatment response of a patient to the application of a particular drug. In the case of livestock a key goal involves predicting variation in production traits, particularly those that are economically important but currently hard to measure. Animals are exceedingly complex, which makes the task of predicting phenotype challenging. The last decade has seen tremendous progress, whereby quantitative genetics theory combined with technical advances in SNP genotyping have allowed statistic models to effectively predict genetic merit. The accuracy of these predictions has been increased through use of dense SNP arrays, and by increasing the number of animals with both genotype and phenotype data in training populations. Further, reductions in sequencing cost have meant it is now feasible to collect the whole genome sequence of hundreds of animals. This is being used to improve genomic prediction, primarily through the utilisation of additional SNP. The availability of whole genome sequence, however, opens much richer opportunities given these datasets directly contain the sequence level differences that control phenotypic variation. Currently functional mutations are indistinguishable from a sea of neutral variation. These difficulties capture the essence of the genome to phenome challenge, which is to successfully interpret genome sequence to predict it's consequence on phenotypic variation.

GENOME ANNOTATION

To meaningfully tackle the genome to phenome challenge, a richly annotated reference genome sequence is required for each of the farm yard animal species. An important milestone on route to this objective was reached in 2014 with the completion of the draft reference genome assembly for sheep (Jiang *et al*. 2014). This completed the collection of available reference genomes for each of the major livestock production species that includes pig (Groenen *et al.* 2012), cattle (Bovine Genome Sequencing and Analysis Consortium 2009), chicken (Burt 2005) and goat (Dong *et al.* 2013). While improvements to these assemblies are ongoing the task of annotating each genome has commenced. Annotation describes the process by which particular sequence characteristics and functional elements are identified in the genome. To date, the features annotated in detail extend only to variation (e.g. SNP and various repeat classes), protein coding genes (intron and exon location) and some aspects of gross sequence classification such as GC content. What is almost completely missing from animal genomes is the accurate identification and annotation of the gene regulatory machinery. The ENCODE project sought to rectify this in human by cataloging the full complement of gene transcripts, their isoforms and the hundreds of thousands of enhancers, transcription factor binding sites and promoter regions active across different cell types (ENCODE Project Consortium 2012). This large and costly international research effort has provided key advancements in our understanding of biology. For example, of the approximately 25,000 human protein coding genes only about 50% are expressed in any given cell type (Romanoski and Glass 2015). Further, it appears possible to identify the combinations of transcription factors responsible for directing the specialisation of precursor cells to differentiate into particular cell types. These fundamental observations represent the first steps towards a more sophisticated ability to understand how DNA sequence and gene regulation together serve to control complex traits.

FUNCTIONAL ANNOTATION OF ANIMAL GENOMES (FAANG)

Animal scientists with the shared goal of producing genome wide maps of functional elements held a planning workshop in January 2014 the Plant and Animal Genome Conference (PAG XXII). The meeting conceptualised the creation of a consortium to coordinate and execute the FAANG project. Subsequent discussion has defined the structure of the FAANG Consortium (i.e. working groups and their roles) and aspects of the FAANG Project (i.e. the operational plan for the science). Key aspects of both are described here, however additional considerations relating to the creation of a common data infrastructure, a centralised data analysis centre, pre-publication data release and the operational principles for participating scientists is available at the consortium website (http://www.animalgenome.org/community/FAANG/). In addition, the consortium recently published a white paper that describes the rationale for the science while providing details about the objectives (The FAANG Consortium 2015). It is important to note that any interested scientists are welcome to participate, and this can be initiated by signing up to the consortium on the website.

SPECIES AND DATA TYPES

Given the ENCODE project focussed on a single species (human) and cost in excess of $150 million dollars, the livestock community recognised early that clear prioritisation was needed to design a project broadly in line with the vastly diminished financial resources likely available to animal scientists. This planning sought to take advantage of i) the declining cost of the sequencing and ii) key lessons from ENCODE relating to the choice of core data and tissue types for investigation. A prerequisite for inclusion in the FAANG project is the availability of a draft reference genome assembly of sufficient quality to serve as the template for annotation. At present, this means the project is confined to cattle, sheep, chicken and pig however additional species are

likely to be included as their genome assemblies improve (e.g. salmon, goat and horse). The consortium has also defined the following set of core assays to be deployed in each species:

i) RNA Transcriptomics
Annotation requires detailed knowledge of the gene transcripts that are present within tissues, along with details of the transcriptional complexity many genes exhibit (e.g. tissue specific isoforms). RNA-sequencing will therefore be used to generate the transcriptome of each core tissue, from each species.

ii) Histone Modification Marks
To map the genomic location of putative promoters, enhancers and transcription start sites (TSSs), CHiP-seq assays will be used that identify DNA sequences that bind to modified histones. The project has prioritised four histone modification marks found to be most informative by the ENCODE projects. These are:

- H3K4me3 that correlates with promoters and transcription start sites
- H3K27me3 which marks silenced genes
- H3K27ac that indicates active regulatory elements
- H3K4me1 which is associated with enhancers and enriched downstream of TTSs

iii) Chromatin Accessibility and Architecture
To complement ii), methods are available that identify 'protein bound' DNA sequences due to chromatin accessibility and architecture. DNaseI footprinting was the first generation of such approaches, however more robust and sensitive approaches have been developed. One is ATAC-seq, and will be used to identify open chromatin. Importantly, the results will be co-analysed with histone modification information to decipher the location of specific protein-DNA binding events to base-pair resolution.

Beyond each of these core assays, the consortium has identified an additional set of data types considered non-essential but informative. These include DNA methylation, antibody dependant direct identification of transcription factor binding sites and genome conformation assays using methods such as Hi-C. Additional detail on all of these assays can be found elsewhere (Lane *et al.* 2014; The Mouse ENCODE Consortium 2014; The FAANG Consortium 2015).

AUSTRALIA'S CONTRIBUTION

To initiate an Australian contribution into the international FAANG consortium, the co-authors have commenced informal discussions to i) collate existing projects that might be included and ii) define the objectives for future projects and strategise how they might be funded. Table 1 shows details of four projects in cattle and sheep, and in each case the focus is largely on transcriptomics. The first two are underway, the third is pending grant approval and the final project is funded and data generation is likely to commence in the last quarter of 2015. It is worthwhile noting very large surveys of genomic variation (SNP and indels), identified by whole genome sequencing projects, will be an important dataset used by the FAANG data analysis teams. In the case of cattle and sheep, these projects are being lead by Australian researchers (e.g. Daetwyler *et al.* 2014). Currently no data generation is planned for *Bos indicus* by non-Australian FAANG partners. Given their importance to the Australian cattle industry, the co-authors have prioritised *Bos indicus* as the focus for joint project applications. Interested parties who would like to become involved should contact any of the authors.

Table 1. FAANG compliant datasets currently being generated by Australian Scientists

Species	Breed	Tissues (n) [1]	Assays	Status	Contacts
Cattle	Holstein	Various(38)	RNA-seq, CHiP-seq	Ongoing	Hayes, Chamberlain
Sheep	Various	GIT[1] (7)	RNA-seq of mRNA (coding only)	Ongoing	Dalrymple, Oddy
Sheep	Various	GIT[1] (7)	RNA-seq of lncRNA, microRNA	App. Pending	Dalrymple, Oddy
Sheep	Rambouillet	Various(20+)	PacBio Iso-seq	Funded	Kijas, Cockett

[1]The number of tissues collected is given in parenthesis, however in some cases only a subset will be used for data generation. GIT is an abbreviation for gastrointestinal tract.

IMPACT FOR ANIMAL SCIENCE

The completed FAANG project will provide a comprehensive data resource describing gene regulation and the genomic elements responsible. The manner in which this resource is used is likely to evolve over time. In the short term, the availability of a genome atlas of regulatory elements should greatly assist the interpretation of whole genome sequencing studies that aim to identify functional variants. Currently the hunt for functional variants is most often successful where a protein coding mutation is responsible, simply because the annotation of animal genomes is most complete for protein coding genes. Conversely regulatory mutations that underpin trait variation are far more difficult to identify, however they may be the most common. FAANG data should greatly assist in elucidating the consequence of variants that directly impact phenotype via alterations in gene action. In the field of genomic prediction, the outcomes from FAANG may propel the field beyond the use of SNP as the sole molecular input data-type for prediction. For example, it may be possible that transcription factor binding site networks, co-expressed gene sets or combinations of these along with SNP genotypes will become the input data that returns higher prediction accuracies than currently available for complex traits. In the short term FAANG data can be used to better annotate SNP and prioritise those likely to directly impact phenotypic variation for exploitation in reduced size SNP panels diagnostic of key traits. Regarding evolutionary questions, the availability of standardised data across at least four species (two ruminants, one monogastric mammal and one bird) should facilitate discoveries concerning those components of the gene regulatory machinery that are conserved and those that are lineage specific. We anticipate a completed FAANG project should provide a general resource for research into mammalian biology and variation in complex traits.

ACKNOWLEDGEMENTS

The FAANG consortium was facilitated by the EC-US Biotechnology Research Task Force.

REFERENCES

Bovine Genome Sequencing and Analysis Consortium (2009) *Science* **324:** 522.
Burt D. (2005) *Genome Res*. **15:** 1692.
Daetwyler H., Capitan A., Pausch H., Stothard P., *et al*. (2014) *Nat. Genet.* **46**: 858.
Dong Y., Xie M., Jiang Y., Xiao N., Du X., et al. (2013) *Nat. Biotechnol.* **31:** 135.
Groenen M., Archiblad A., Uenishi H., Tuggle C., Takeuchi Y. *et al*. (2012) *Nature* 491: 393.
Jiang Y., Xie M., Chen W., Talbot R., Maddox J.F. *et al*. (2014) *Science* **344**: 1168.
Lane M. Niederhuth C.E., Lexiang J., Schmitz R.J. (2014) *Annu.Rev.Genet.***48**:49.
Romanoski C. and Glass C. (2015) *Nature* **518:** 7539.
The ENCODE Project Consortium (2012) *Nature* **489:** 57.
The Mouse ENCODE Consortium. (2014) *Nature* **515**: 355.
The FAANG Consortium. (2015) *Genome Biol*. **16:**57.