

**RELIABILITIES OF AUSTRALIAN DAIRY GENOMIC BREEDING VALUES
INCREASE THROUGH THE ADDITION OF GENOTYPED FEMALES WITH
EXCELLENT PHENOTYPES**

**J.E. Pryce^{1,2}, P. Douglas³, C.M. Reich¹, A.J. Chamberlain¹, P.J. Bowman^{1,2}, T.T.T. Nguyen¹,
B.A. Mason¹, C.P. Prowse-Wilkins¹, G.J. Nieuwhof^{1,4}, T. Hancock^{1,4}, M. Shaffer⁴ and B.J.
Hayes^{1,5}**

¹ Department of Economic Development, Bundoora, VIC 3083, Australia

² La Trobe University, Bundoora, VIC 3083, Australia

³ Holstein Australia, Hawthorn, VIC 3122, Australia

⁴ DataGene, Bundoora, VIC 3083, Australia

⁵ University of Queensland, Brisbane, QLD, Australia

SUMMARY

Ginfo is a large-scale genotyping project to increase the size of the Australian dairy reference population. In total, there were 32,386 cows from 103 herds with excellent records located across Australia's main dairy regions. The increase in the reliabilities of breeding values for young genomic bulls (without daughters) was between 5% and 7% in Holsteins and between 2% and 3% in Jerseys. For example, in Holsteins, the reliability of daughter fertility breeding values increased from 41% to 46%. Incorporating genotypes from herds selected on the quality of their phenotypes has increased the reliability of genomic breeding values.

INTRODUCTION

Phenotypic data underpins the calculation of both traditional and genomic breeding values. A reference population of genotyped individuals with phenotypes is required to calculate associations between genetic markers and phenotypes and form a genomic prediction equation. Without sufficient data, the relationship between the reference population and the general population weakens and so does the relevance of the genomic prediction equations.

In Australia, the male reference populations comprise around 4000 Holsteins and 1000 Jerseys. Previous research investments have already resulted in female populations of approximately 10,000 Holstein 4000 Jersey females being added to the national reference populations of the respective breeds. These data were from projects that focused on cows with large quantities of phenotypes. Instead, the aim of Ginfo was to select herds that had high quality phenotypes.

The aim of this study was to quantify the change in reliability of genomic breeding values for Australian breeding values through adding the Ginfo population to the reference population.

MATERIALS AND METHODS

Herd Selection. To qualify for the reference population, known as Ginfo (Genomic Information Nucleus), Australian dairy herds were scored according to the quality of the records contributing to the national database using an index that rewards cows with fertility, conformation, survival, workability, somatic cell count and milk yield data; in the scoring system, the maximum score was 25 and having complete fertility phenotypes can make up 10 of these points. The highest scoring herds (n=103) were invited to participate in the project.

The 103 Ginfo herds have been contributing records on 32,386 daughters of 2,917 bulls to the Ginfo project. Tail hair samples were collected from all the cows in the recruited herds for genotyping and data on milk production traits, somatic cell count, mating, pregnancy and calving data for multiple parities were provided to DataGene. First parity cows from Ginfo herds were also type classified by Holstein Australia. The herds were from across Australia's main dairy regions

Dairy

with an intention to proportionally represent dairy cow populations. For example, two thirds of the herds were in Victoria aligning with the distribution of dairy cows across Australia.

Genotyping by sequencing (GBS). Genotyping by sequencing (GBS) was used to procure genotypes. The GBS methodology used has previously been described by Chamberlain *et al.* (2015). Briefly, probes were designed to the flanking sequencing of 9,102 target SNP, of which 5,119 were part of the Illumina Infinium Bovine SNP50 beadchip. The HiSeq2000 and HiSeq3000 genome analyser platforms using single read chemistry were used for sequencing.

Quality control steps of sequence reads were as follows: 1) poor quality bases (qscore <20) were removed using scripts developed in house; 2) alignment was with BWA v0.7.7; 3) Samtools v0.1.19 (Li *et al.*, 2009) mpileup tool was used to create vcf files and allele counts at the 9,102 target SNPs and 4) Allele counts were used to call genotypes, where the total count was ≥ 6 and a heterozygote had a minor allele frequency > 0.167 . The genotypes in UMD 3.1 forward format were converted to Illumina's top-top format. The next step was imputation of GBS genotypes to those used by DataGene in routine genomic evaluations (Nieuwhof *et al.*, 2010).

All animals were imputed to a 50K evaluation panel using Fimpute (Sargolzaei *et al.*, 2014). The Ginfo project also enhanced the DataGene evaluation SNP panel to include new variants which were identified by whole genome sequence (WGS) analysis which were found to be located near new QTLs for the traits within the evaluation. These WGS variants were added to the DataGene evaluation panel through the Ginfo GBS genotypes and all other animals were imputed for these WGS SNPs.

Impact of Ginfo population on reliability of genomic selection. The Ginfo cows and their associated phenotypes were added to the genomic reference population. In April 2016, when our comparisons were done, the existing reference populations comprised 4,172 bulls and 10,254 cows for Holsteins and 1,097 bulls and 4,232 cows for Jerseys. The cows that were already included in the reference population were selected using similar selection criteria for phenotype quality, as described already for Ginfo.

Reliabilities were estimated for all traits evaluated by DataGene using software developed in house for genomic selection (Nieuwhof *et al.*, 2010) implementing the mixed model equations for genomic selection as described by Garrick (2007). The reliabilities of genomic bulls with no daughters were compared when estimated with and without Ginfo cows in the reference population.

RESULTS AND DISCUSSION

The number of Ginfo cows added to the reference population was 17,108 and 3,347 for Holsteins and Jerseys respectively. At the time the Ginfo population was added to the Australian national reference population, they represented 54% and 39% of the Holstein and Jersey populations respectively.

On average the increase in reliability from adding Ginfo to the reference population was 5.8% and 2.5% for young genotyped Holstein and Jersey bulls respectively (Figure 1). The impact varied by trait, with gains of between 5% and 7% for Holsteins and between 2% and 3% for Jerseys. For example, in Holsteins, the reliability of daughter fertility increased from 41% to 46%, while overall type increased from 42% to 49%. This is similar to approximations derived by applying the equation of Daetwyler *et al.* (2008) to predict the reliability of genomic prediction for varying reference population sizes. The scores for herds that are in Ginfo are on average higher for Holsteins than Jerseys, which is a consequence of the relative population sizes. This could have partly contributed to the smaller increase in reliability for Jerseys compared to Holsteins.

One of the main questions in the design of future reference populations is whether to focus on increasing reliabilities through genotyped bulls with large progeny groups, or on genotyped cows with their own phenotypes (Gonzalez-Recio *et al.*, 2014; Chesnais *et al.*, 2016). Another

alternative, discussed by Plieschke *et al.* (2016) is genotyping and phenotyping a fixed number of first crop daughters, as this increases the reliability of the sire. The general conclusions of Chesnais *et al.* (2016) are that when phenotypes are inexpensive and easy to measure on a large scale for key traits of interest, bull reference populations are better, while for expensive or difficult to measure traits, it is preferable to have a reference population of genotyped females. However, there is also a case for female reference populations, where the usual source of new phenotypes (i.e. the number of progeny-tested bulls) is in decline. In Australia, the number of bulls with sufficient daughters with publishable proofs for production traits by year of birth has gradually been declining, by around 60 per year. Consequently, a genomic reference population that does not solely rely on progeny-tested sires is important. When large male reference populations are already available, the impact of adding females on reliabilities is comparatively small; so the value for these sorts of reference populations is more around the new traits that can be measured in dedicated reference populations.

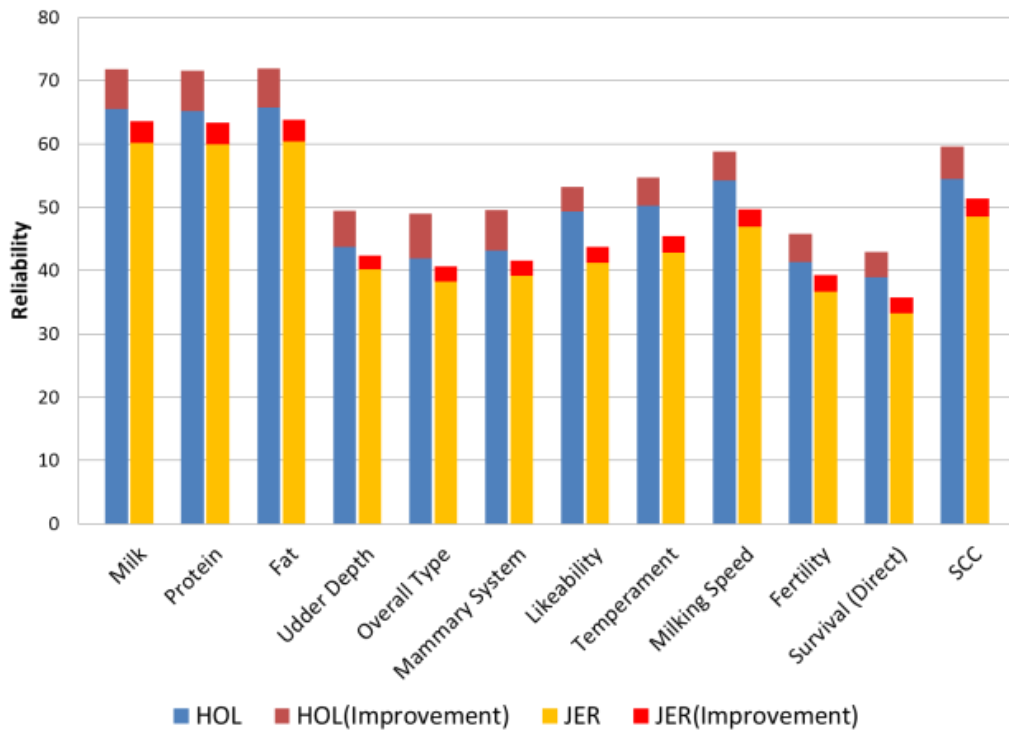


Figure 1. Reliabilities of traits with and without the Ginfo population

The Ginfo reference population is projected to encompass approximately 60,000 milking animals in 200 herds to reflect the genetics, location and farm systems in the broader Australian dairy population. Ginfo is anticipated to become a primary source for the Australian industry’s ongoing evaluation of the current suite of genomic Breeding Values. In addition we are also investigating the collection of emerging and new phenotypes of interest to farmers particularly for animal health traits and traits associated with resource availability and efficiency (Abdelsayed *et al.*, 2017).

Dairy

One of the philosophies in establishing Ginfo was to develop relationships with Australian dairy farmers who have a shared interest in the value of high quality phenotypes and genotype results. Although the model we have used to date included all genotyping costs being covered through research funding, we envisage that this will change as we move to a model where farmers pay for a much larger proportion of the genotyping cost themselves. While the genotyping results (breeding values) of lactating cows may have limited use for decision making, there is considerable value in genotyping results for heifers, most notably in selecting the best replacements (Pryce and Hayes, 2012 Calus *et al*, 2013). Therefore, the investment strategy needs to balance the benefits to the farmer versus the benefits to the broader dairy industry.

ACKNOWLEDGEMENTS

The Ginfo research project was funded by Dairy Futures CRC (Melbourne, Australia). Ginfo is now an industry initiative and is part of DataGene (Melbourne, Australia).

REFERENCES

- Abdelsayed, M., Haile-Mariam M and Pryce, J.E. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22** (submitted)
- Calus, M.P.L., Bijma, P., van Arendonk, J. and Veerkamp, R.F. (2013) *Ann. Mtg Europ Assoc. Anim. Prod.* **26**.
- Chamberlain A.J., Beatson P., Mason B.A., Prowse-Wilkins C, Reich CM., Schrooten C, Hayes B.J., Tibbits J., and M. Hayden (2015) *Proc. Assoc. Advmt. Anim. Genet.* **21**: 69-72.
- Chesnais, J.P., Cooper, T.A., Wiggans, G.R., Sargolzaei, M., Pryce, J.E. and Miglior, F. (2016) *J. Dairy Sci.* **99**: 2413-2427.
- Daetwyler, H.D., Villanueva B. and Woolliams J.A., (2008). *PLoS One*, **3**: e3395.
- Garrick, D. J. (2007) *J. Anim. Sci.* **85**: 376.
- Gonzalez-Recio, O., Coffey, M.P. and Pryce, J.E. (2014). *J. Dairy Sci.* **97**: 7905-7915.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. and Durbin, R. (2009) *Bioinformat.* **25**: 2078-2079.
- Nieuwhof, G.J., Beard, K.T., Konstantinov, K.V., Bowman, P.J., and Hayes, B.J. (2010) *Proceedings of 2010 Interbull meeting, Riga, Latvia*
- Plieschke, L., Edel, C., Pimentel, E.C., Emmerling, R., Bennewitz, J. and Götz, K.U. (2016) *Genet. Sel. Evol* **48**: 73.
- Pryce, J.E. and Hayes, B.J. (2012). *Anim. Prod.Sci.* **52**: 180-184.
- Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2014). *BMC genomics*, **15**: 478.