Proc. Assoc. Advmt. Anim. Breed. Genet. 19:331-334

USE OF GENOMIC INFORMATION TO ESTIMATE BREEDING VALUES FOR CARCASS TRAITS IN SHEEP

A.A. Swan[1], D.J. Brown[1], B. Tier[1], 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 Rural and Environmental Science, University of New England, Armidale, NSW 2351
[3]CRC for Sheep Industry Innovation, Armidale, NSW 2351

SUMMARY

Breeding values for carcass traits were estimated in a multi-breed sheep population using phenotypic, pedigree, and genomic information. This was achieved by incorporating a genomic relationship matrix into the standard pedigree based relationship matrix used in an animal model genetic evaluation. Heritability estimates were generally very close to estimates from a model using pedigree information only. A group of young rams genotyped but not measured for the traits in question were included in the analysis, and the accuracy of their estimated breeding values estimated using the prediction error variances of the fitted model increased by between 14 and 24 percentage points when genomic information was used. However, these accuracies were between 12 and 24 percentage points higher than observed accuracies, indicating that the scaling of the genomic relationship matrix was incorrect. Further research is required on the implementation of the method in multi-breed data.

INTRODUCTION

Genomic selection using information from high density SNP marker panels can improve the accuracy of selection considerably, depending on the context. Van der Werf (2009) showed that genomic selection in sheep could increase selection response in overall merit by 30 to 40%, with the impact being greatest for traits which are not routinely measured on young breeding animals such as carcass, adult wool traits, female fertility, and disease traits.

When commercially relevant animals are genotyped, the benefits for breeding programs will be best captured by incorporating this genomic information into estimated breeding values (EBVs). The challenge for implementation is how to deal with a mixture of animals with records on important traits that may or may not have been genotyped. Two approaches are possible, with the first being a multi-step process where an association analysis is performed to estimate genomic breeding values (GBVs) for animals with genotypes, with these GBVs then either included in a genetic evaluation model as additional traits (Johnston *et al* 2009) or blended with EBVs from an existing genetic evaluation using selection index theory (e.g. Harris and Johnson 2010). The second and preferred approach is to simultaneously include all genomic, phenotypic, and pedigree information in a single analysis. Such an approach has been developed by Aguilar *et al.* (2009), and in this paper we implement this method to estimate breeding values enhanced by genomic information for carcass traits in sheep.

MATERIALS AND METHODS

Data used were obtained from the Sheep CRC's Information Nucleus Flock (INF) (Fogarty *et al.* 2007). This is a multi-breed population, with approximately 100 industry sires from terminal, maternal and Merino sires mated annually to Merino and crossbred dams at eight sites across Australia. The progeny are measured for a wide range of traits, including the carcass and meat

· AGBU is a joint venture of NSW Dept. of Industry and Investment and the University of New England

quality traits used in this study and a proportion are genotyped using a high density 50K SNP marker panel.

The traits considered in this study were hot carcass weight (hcwt, kg), carcass eye muscle depth (cemd, mm), carcass fat depth at the C site (ccfat, mm), lean meat yield (lmy, %), shear force at day 5 post slaughter (shf5, Newtons), and carcass intramuscular fat (cimf, %). A summary of the data is shown in Table 1. Animals were measured in 2008, 2009 and 2010 for some traits, with the number of animals with records ranging from 3554 to 6710, and between 2711 and 3668 genotyped. Mean age at slaughter was 262 days. There were between 179 and 313 sires, and 155 to 209 of these also had genotypes. In addition, 249 young industry rams with genotypes only were included in the analysis. These young rams were distributed across the main breeds in the data.

Table 1. Data summary for carcass traits analysed (see text for trait definitions)

	hcwt	cemd	ccfat	lmy	shf5	cimf
Records	6710	5760	5611	4789	3554	3762
Records genotyped	3668	3590	3478	2121	2711	2860
Mean	22.9	30.1	4.1	58.0	26.5	4.4
Sires	313	312	311	312	179	184
Sires genotyped	209	208	208	208	155	160

Single trait models were used as follows:

$$y = X\beta + ZQg + Zu + e$$

Where y is the vector of records, $X\beta$ represents fixed effects, ZQg represents breed effects, Zu breeding values, and e random residual effects. The fixed effect common to all traits was contemporary group defined in sub-classes of year of birth, site, management group, kill date. For shear force, an additional sub-class for test laboratory was also included. Other effects included age of dam (hcwt), age of measurement (hcwt, lmy), birth type (hcwt), rearing type (hcwt), and hcwt (cemd, ccfat, shf5, and cimf).

Breed effects were fitted as partial regressions of performance on the proportion of genes from each breed, with the matrix Q containing breed proportion coefficients for each animal in the pedigree for analysis animals. These were derived from a pedigree merged across all of the separate genetic evaluation analyses performed in Australia, and in theory giving the best available information on breed composition. There were 29 breeds represented in the data, with Merinos sub-divided into ultrafine, fine-medium, and strong wool strains. Several breeds were not well represented, and to reduce problems with estimability breeds were fitted as random effects.

Breeding values were estimated using two methods. In the first (AEBV), a standard animal model was fitted using the numerator relationship matrix (A) for all animals in the pedigree. This pedigree was constructed to include two generations of ancestral pedigree for the animals with records and the young industry rams with genotypes, and included 17,195 animals in total. Hence, for this model $var(u) = A.\sigma_a^2$ where σ_a^2 is the additive genetic variance.

In the second method (HEBV), the inverse of the numerator relationship matrix A was replaced by the following matrix as derived by Aguilar *et al.* (2009):

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

Where A is the numerator relationship matrix for the entire pedigree as before, G is a genomic relationship matrix for the subset of genotyped animals, and A_{22} is the sub-partition of the numerator relationship matrix for those animals. Firstly, a raw genomic relationship matrix (G_r) was calculated from the 50K SNP genotypes following VanRaden (2008), scaled so that the average diagonal element was 1. Then G was calculated as $\lambda G_r + (1-\lambda)A_{22}$using $\lambda = 0.95$ as the weighting factor as proposed by Aguilar *et al.* (2009) and Forni *et al.* (2011) to alleviate problems with singularities in the genomic relationship matrix.

The models were run in ASReml (Gilmour *et al.* 2009), with H^{-1} included as a user specified matrix, and were allowed to converge to REML estimates of the variance components. Estimated breeding values and accuracies from the two methods were then compared within genotyped and un-genotyped progeny and sires, and for the young industry rams with genotypes. While breeding values for total genetic merit would normally be estimated as $Q\hat{g} + \hat{u}$ where $\hat{g}$ and $\hat{u}$ are the estimates of breed effects and breeding values, comparisons were based only on $\hat{u}$, as it was the component directly affected by genomic information. Accuracies were calculated as $\sqrt{1 - PEV_i/(d_{ii} \times \hat{\sigma}_a^2)}$ % where PEV_i is the prediction error variance for the i^{th} animal obtained from ASReml output, d_{ii} is the diagonal element of either H or A for the i^{th} animal, and $\hat{\sigma}_a^2$ is the estimated genetic variance. They were compared with observed accuracies calculated independently of this study by splitting the data for progeny with genotypes into training and validation sets, calculating a genomic prediction equation in the training set, and then evaluating its accuracy in the validation set (H.D. Daetwyler, pers. comm.).

RESULTS

Single trait estimates of parameters for the AEBV and HEBV models are shown in Table 2. Heritability estimates for the two methods were similar for cemd, ccfat, lmy and shf5. For hcwt heritability was 0.13 lower for HEBV, while for cimf it was 0.05 higher.

Table 2. Parameter estimates for heritability (h^2), additive genetic variance (σ_a^2), phenotypic variance (σ_p^2) and between breed variance (σ_{gg}^2) for the AEBV and HEBV methods

Param.	Method	hcwt	cemd	ccfat	lmy	shearf5	cimf
h^2	AEBV	0.55 ± 0.04	0.31 ± 0.04	0.28 ± 0.04	0.35 ± 0.04	0.30 ± 0.05	0.43 ± 0.06
	HEBV	0.42 ± 0.03	0.30 ± 0.03	0.27 ± 0.03	0.36 ± 0.04	0.31 ± 0.04	0.48 ± 0.04
σ_a^2	AEBV	3.22 ± 0.30	2.34 ± 0.33	0.84 ± 0.12	2.13 ± 0.29	14.03 ± 2.52	0.28 ± 0.04
	HEBV	2.41 ± 0.22	2.30 ± 0.28	0.82 ± 0.10	2.21 ± 0.26	14.71 ± 2.25	0.32 ± 0.04
σ_p^2	AEBV	5.83 ± 0.13	7.51 ± 0.16	3.00 ± 0.06	6.11 ± 0.14	46.88 ± 1.24	0.66 ± 0.02
	HEBV	5.74 ± 0.12	7.59 ± 0.16	3.02 ± 0.06	6.21 ± 0.14	47.50 ± 1.28	0.68 ± 0.02
σ_{gg}^2	AEBV	14.13 ± 4.87	3.18 ± 1.39	1.06 ± 0.48	3.38 ± 1.38	5.54 ± 3.35	0.20 ± 0.11
	HEBV	13.06 ± 4.49	2.75 ± 1.29	0.91 ± 0.45	2.61 ± 1.19	6.04 ± 4.19	0.15 ± 0.10

Including genomic information had a small impact on breeding values of measured progeny and their sires. Correlations between AEBV and HEBV estimated breeding values averaged 0.98 and 0.94 for un-genotyped progeny and sires respectively, 0.94 and 0.89 for genotyped progeny and sires, and 0.48 for young rams. For traits where heritability showed little change between methods, accuracies for progeny and sires were similar, while for hcwt the lower heritability led to a reduction in accuracy, and for cimf the higher heritability lead to an increase in accuracy.

Results for young rams that were genotyped but not measured are shown in Table 3. Mean accuracies for HEBV ranged from 32 to 40%. These means represented a mean improvement of between 14 and 24 percentage points over AEBV accuracies for these animals. However, HEBV accuracies were ranged from 12 to 24 percentage points higher than observed accuracies.

Table 3. mean accuracy (%) of HEBV for young rams, increase in accuracy (Δ = HEBV – AEBV) for young rams, and observed accuracy (H.D. Daetwyler, pers. comm.)

	hcwt	cemd	ccfat	lmy	shf5	cimf
Accuracy	40	37	36	37	32	37
Accuracy Δ	23	14	14	16	20	24
Observed accuracy	27	25	12	21	8	19

DISCUSSION

One of the challenges with the HEBV method is to ensure that the genomic relationship matrix is scaled appropriately so that it is compatible with the pedigree based relationship matrix in H. Incorrect scaling can lead to inflated estimates of genetic variance and accuracies of breeding values (Forni *et al.* 2011). Use of a normalised G in this study as proposed by Forni *et al.* should lead to similar estimates of genetic variance for both methods but with lower standard errors for HEBV. The results presented in Table 2 were generally consistent with this expectation. However, the disparity between accuracies calculated from the HEBV method and observed accuracies in Table 3 indicates that there was a problem with the scaling of G. In a subsequent analysis using data only from Merinos HEBV accuracies were not inflated relative to the observed accuracies. This suggests that the problem is due to the multi-breed nature of the data.

While the ability of the HEBV method to simultaneously use all records together with pedigree and genomic information has obvious advantages, further research is needed on its application in multi-breed data.

ACKNOWLEDGEMENTS

This work was funded by MLA Project B.SGN.0127 and the CRC for Sheep Industry Innovation. We also thank Hans Daetwyler for his assistance with the genomic data and analyses.

REFERENCES

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. and Lawlor T.J. (2009) *J. Dairy Sci.* **93**:743.

Fogarty N.M., Banks R.G., Van der Werf J.H.J., Ball A.J. and Gibson J.P. (2007) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17:**29.

Forni S., Aguilar I. and Misztal I. (2011) *Genet. Sel. Evol.* **43**:1.

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

Harris B.L. and Johnson D.L. (2010) *J. Dairy Sci.* **93**:1243.

Johnston D.J., Tier, B. and Graser H.-U. (2009) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **18**: 30.

Van der Werf J.H.J. (2009) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **18**:38.

VanRaden P.M. (2008) *J. Dairy Sci.* **91**: 4414.