Proc. Assoc. Advmt. Anim. Breed. Genet. 20:220-223

ACCURACY OF GENOMIC PREDICTION FROM MULTI-BREED SHEEP REFERENCE POPULATION

N. Moghaddar[1, 2, 4], 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

Genomic estimated breeding values (GEBV) were calculated based on a combination of purebred and crossbred sheep for birth weight, weaning weight and post weaning weight using genomic best linear unbiased prediction (GBLUP). The genomic relationship matrix (G) was calculated based on population wide or breed of haplotype specific allele frequency using the 50k ovine Illumina SNP-chip. The accuracy of genomic prediction was estimated based on the correlation between genomic breeding value and an accurate breeding value based on progeny records. The result showed better genomic prediction accuracy for breeds with higher representation in the combined reference populations. Accuracies slightly decreased when the reference set contained a significant set of additional animals from another breed. This study showed no extra accuracy from across breed information using 50k SNP marker panel. The result showed a small but non-significant increase in accuracy when using breed specific allele frequencies in the calculation of G.

INTRODUCTION

Genomic selection can significantly increase the rate of genetic progress in quantitative traits by providing extra accuracy from exploiting Mendelian sampling and by reducing the generation interval (Meuwisen *et al*. 2001; Schaeffer 2006; Banks *et al*. 2009; Dalton 2009; van der Werf 2010). The size of the reference population has an important impact on the accuracy of genomic prediction (Goddard 2009). In the sheep industry data are often available from a mixture of breeds, multiple strains within a specific breed or from crossbreds. Combining populations of different pure and crossbred animals would be an advantage if it could be shown to increase the accuracy of genomic prediction, particularly for breeds which are not well represented in the combined reference population. The objective of this study is to assess the effect of a combined sheep reference population on accuracy of within breed genomic prediction using real data. The accuracy of genomic prediction was compared between GEBV prediction from purebred, crossbred and a combination of purebred and crossbred data which was extracted from a large multi breed/crossbred sheep reference population. Furthermore, two strategies in calculating the genomic relationship matrix (G) were compared to investigate the effect of accounting for different marker allele frequencies between breeds.

MATERIALS AND METHODS

Reference population and phenotypic data. The reference populations tested contained either purebred Merino sheep (M) or crossbreds of Border Leicester and Merino (BLxM), or a combination of both. Three population sizes (1000, 2000 and 3000) were used for the purebred Merino reference sets and these were compared with 3 sets where the purebred Merino populations were augmented with 1472 BLxM crossbreds. Data was extracted from the Sheep CRC Information Nucleus database (van der Werf *et al.* 2010). The traits investigated were birth weight

(BW), weaning weight (WW) and post weaning weight (PWW). Phenotypic means and standard deviations were 4.76 ± 1.02, 25.4 ± 5.78 and 40.2 ± 8.2 respectively.

Genotypes and validation population. Animals were genotyped using the 50K Ovine chip (Illumina Inc., SanDiego, CA, USA). This chip provided 48,559 SNP genotypes after applying quality control. The accuracy of GEBV was estimated as the correlation of GEBV and accurate EBV based on pedigree and phenotypes in an independent group of animals which had been genotyped for use as a validation population. The validation population comprised 175 Merino sires and 55 Border Leicester sires with average EBVs accuracies of 0.92 and 0.98, respectively. Comparison of correlation coefficient of two dependent samples was used as test statistics.

Statistical methods. Genomic best linear unbiased prediction (GBLUP) was used to calculate the GEBV using ASReml (Gilmour 2009). The following model was used for analysis of data: $y = Xb + Z_1g + Z_1Qq + Z_2m + e$ where y is a vector of phenotypes, b is a vector with fixed effects, g is the random additive genetic effect of the animal, q is a vector with random breed effects, m is a vector with maternal effects, and e is vector of random residual effects, X and Z_1 and Z_2 are incidence matrices and Q contains breed proportions as derived from a deep pedigree. g, q and e are considered normally distributed as $g \sim N(0, G\sigma_g^2)$, $q \sim N(0, Q\sigma_q^2)$, and $e \sim N(0, I\sigma_e^2)$, respectively, where G is the genomic relationship matrix. 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. G was calculated using two approaches according to VanRaden (2008). In one approach G was calculated using the overall marker allele frequencies of the entire population (G1) while in the second approach the breed of haplotype specific marker allele frequencies were used (G2).

Table 1. Accuracy of genomic prediction from different reference populations for birth weight (BW) for Merino and Border Leicester (BL)

Reference population		Breed proportion (%)		GEBV accuracy[1]			
				G1		G2	
Type	Size	BL	Merino	BL	Merino	BL	Merino
(1) = Merino	1000	0.0	100	-0.03 [b]	0.38 [bc]	-0.03 [b]	0.38 [bc]
(2) = Merino	2000	0.0	100	-0.10 [ab]	0.42 [cd]	-0.10 [ab]	0.42 [cd]
(3) = Merino	3000	0.0	100	-0.16 [a]	0.47 [d]	-0.14 [a]	0.47 [d]
BLxMerino	1472	50.7	47.2	0.24 [c]	0.29 [a]	0.24 [c]	0.29 [a]
BLxMerino + (1)	2472	30.1	68.3	0.23 [c]	0.36 [b]	0.24 [c]	0.39 [bc]
BLxMerino + (2)	3472	21.4	77.3	0.17 [c]	0.39 [bc]	0.17 [c]	0.39 [bc]
BLxMerino + (3)	4472	16.6	82.4	0.18 [c]	0.42 [cd]	0.18 [c]	0.42 [cd]

G1: Genomic relationship matrix based on all SNP allele frequency. G2: Genomic relationship matrix based on haplotype SNP allele frequency. Different superscripts for accuracies indicate statistical differences.

RESULTS AND DISCUSSION

Tables 1, 2 and 3 show the GEBV accuracy for BW, WW and PWW according to the two methods to calculate G respectively. The results show that the GEBV accuracy for Merino sheep

increased as the size of the reference population increases, both for purebred and combined purebred and crossbred animals. However, the accuracies were higher when prediction was based on purebred Merinos, and accuracy slightly decreased if the BLxM batch of crossbred animals was added to a purebred Merino reference population.

The GEBV accuracy for BL validation animals in all three weight traits is higher when the genomic prediction is based on a population with a maximum proportion of BL animals. Adding additional purebred Merino animals to the BLxM reference population reduces the accuracy for BL validation animals. Differences were not statistically significant for BW but they were significant for WW and PWW ($P<0.05$).The results also show a slightly higher GEBV accuracy in some cases from using breed specific marker allele frequency (G2) compared to overall population marker allele frequency (G1) for construction of the genomic relationship matrix. However, most of these differences were not significant ($P<0.05$).

The results suggest that the genomic prediction accuracy within a specific breed is mainly determined by the effective number of haplotypes of that breed in the reference population. The accuracy of GEBV for Merinos increased based on prediction from a larger reference population. The rate of increase in accuracy as well as the level of accuracy in Merino was lower when prediction was based on a combination of crossbred and purebred Merinos compared to prediction from only purebred Merinos from a similar population size. This indicates neutral to some negative effect of adding BL haplotypes to the reference population. The accuracy of GEBV for BL when predicted from combined BLxM and purebred Merinos decreased with a decreasing proportion of BL haplotypes, indicating a negative effect of Merinos on accuracy of genomic prediction for BL animals. Genomic prediction from BLxM on their own provides some predictive power for Merinos because all progeny used had Merino dams.

Table 2. Accuracy of genomic prediction from different reference population for weaning weight (WW) for Merino and Border Leicester (BL)

Reference population		Breed proportion (%)		GEBV accuracy[1]			
				G1		G2	
Type	**Size**	**BL**	**Merino**	**BL**	**Merino**	**BL**	**Merino**
(1) = Merino	1000	0.0	100	-0.07 ^{b}	0.42 ^{b}	-0.06 ^{b}	0.42 ^{b}
(2) = Merino	2000	0.0	100	-0.13 ^{b}	0.49 ^{c}	-0.13 ^{b}	0.49 ^{c}
(3) = Merino	3000	0.0	100	-0.26 ^{a}	0.51 ^{c}	-0.22 ^{a}	0.51 ^{c}
BL*Merino	1547	50.0	47.6	0.32 ^{d}	0.31 ^{a}	0.32 ^{d}	0.31 ^{a}
BL*Merino + (1)	2547	30.3	67.6	0.22 ^{c}	0.43 ^{b}	0.24 ^{c}	0.41 ^{b}
BL*Merino + (2)	3547	22.1	76.8	0.16 ^{c}	0.46 ^{b}	0.18 ^{c}	0.45 ^{b}
BL*Merino + (3)	4547	17.0	82.3	0.17 ^{c}	0.47 ^{bc}	0.18 ^{c}	0.44 ^{b}

[1] As defined in Table 1

The degree of relationship between validation and reference population animals affects the accuracy of genomic prediction (Habier *et al.* 2007) and therefore genomic relationships between reference and validation populations were explored. There was on average a low to moderate genomic relationship between Merino validation animals and the reference populations while it

was almost close to zero between BL validation sires and the purebred Merino reference population.

This study showed that the accuracy of GEBV prediction from a multi breed reference population depends highly on breed representation in the reference population, both through numbers and proportion. Daetwyler *et al.* (2010) showed that across breed information does not contribute to genomic prediction accuracy using the 50k marker density. This study showed neutral to negative effect of adding information from animals of a different breed. Applying denser SNP marker panels could potentially lead to better prediction from across breed information. More investigations with larger validation population and also with denser genetic markers are required.

Table 3. Accuracy of genomic prediction from different reference population for post weaning weight (PWW) for Merino and Border Leicester (BL)

Reference population		Breed proportion (%)		GEBV accuracy			
				G1		G2	
Type	**Size**	**BL**	**Merino**	**BL**	**Merino**	**BL**	**Merino**
(1) = Merino	1000	0.0	100	-0.02 [a]	0.53 [b]	0.00 [a]	0.53 [b]
(2) = Merino	2000	0.0	100	-0.04 [a]	0.57 [bc]	-0.04 [a]	0.57 [bc]
(3) = Merino	3000	0.0	100	-0.08 [a]	0.59 [c]	-0.07 [a]	0.59 [c]
BL*Merino	1514	50.7	47.2	0.49 [c]	0.45 [a]	0.49 [b]	0.45 [a]
BL*Merino + (1)	2514	30.5	68.2	0.42 [bc]	0.56 [bc]	0.47 [b]	0.57 [bc]
BL*Merino + (2)	3514	21.8	77.2	0.37 [b]	0.54 [bc]	0.42 [b]	0.57 [bc]
BL*Merino + (3)	4514	17.0	82.3	0.36 [b]	0.56 [bc]	0.41 [b]	0.57 [bc]

G1: Genomic relationship matrix based on overall SNP allele frequency. G2: Genomic relationship matrix based on each population SNP allele frequency.

ACKNOWLEDGEMENTS

The authors are thankful to Daniel Brown (AGBU/UNE) for consultations and providing industry sires ASBVs, Klint Gore for performing quality control and database management and project executive staff at the various research herds. Australian Wool Innovation and Meat and Livestock Australia made the data available from the SGP.

REFERENCES

Banks R.G. and van der Werf J.H.J. (2009) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **18**:430.

Daetwyler H.D., *et al.* (2010) *Animal Production Science* **50:**1004.

Dalton R. (2009). *Nature* **457**: 369.

Gilmour A.R., *et al.* (2009) ASReml 3. ww.vsni.co.uk

Goddard M.E. (2009) *Genetica* **136:**245.

Habier D., *et al.* (2007) *Genetics* **177**:2389.

MeuwissenT., *et al.* (2001) *Genetics **157**:1819*

Schaeffer L.R. (2006) *J. Anim. Breed. Genetic* **123:** 218.

Van der Werf J.H.J. (2010) *Anim.Prod.Science* **50:** 998.

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