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$_{(μ)}$, 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