Genomic estimation of additive and dominance genetic variance and their effect on the accuracy of genomic prediction of sheep

N. Moghaddar[1,2], J.H.J van der Werf[2,1].
[1]Sheep-CRC, Armidale, NSW 2351, Australia, [2]School of Environmental & Rural Sci., University of New England (UNE), Armidale, NSW 2351, Australia,.

ABSTRACT: Additive and dominance variance of weaning weight (WWT) and post weaning weight (PWWT) for purebred Merinos and crossbreds of Merino and other breeds were estimated. Additive and dominance genomic relationships were calculated based on 48,599 SNP marker genotypes. Dominance variation was 3.61% and 5.58% of phenotypic variance for WWT and PWWT, respectively in purebreds and 9.2% and 17.1% for WWT and PWWT in crossbreds. The likelihood of the model was improved by including dominance effects, particularly for crossbred data. The accuracy of within breed genomic breeding value based on prediction from purebreds was similar for additive and additive plus dominance model but showed between 0.3% and 2.1% increase based on prediction from crossbreds and using additive plus dominance model. Fitting both additive and dominance effects of marker genotypes provides either similar or higher GBV accuracy depending on the value of dominance variance.
Keywords: dominance genetic effect; genomic prediction

Introduction

Genomic prediction in animal breeding involves the prediction of breeding value of genotyped selection candidates based on phenotypic and genotypic data from a reference population. Such a reference population could exist of many breeds, especially in beef cattle or sheep breeding, and this could provide some challenges. Moghaddar et al. (2014) showed that genomic prediction for a particular breed could become less accurate when the reference population contains data on crossbred or other breeds. One way to improve prediction from multi-breed reference populations is to account for dominance effects. In animal breeding the effect of dominance, as a non-additive genetic effect, is usually ignored in genetic evaluation of selection candidates, because a reliable estimate of dominance variance from pedigree information needs large family size of particular type (e.g. large full-sib families), and computations are complex (Misztal et al, (1998)). Furthermore, for genetic improvement of quantitative traits we are mainly interested in the additive genetic effect.

The statistical estimation of additive genetic effects and hence breeding values is based on allele substitution effects (Falconer (1981)). Allele substitution effects depend on genotypic values and allele frequencies, and part of the dominance effects are incorporated in the estimates of additive genetic effects, particularly in more extreme allele frequencies (Falconer (1981); Hill et al (2008)).

Genomic prediction is based on estimating effects of genotypes at single nucleotide polymorphism (SNP) marker loci covering whole the genome, with part of them in linkage disequilibrium with quantitative trait loci (QTL). Because of our knowledge on SNP genotypes, we have a much better handle on estimating dominance effects compared to using pedigree relationships. This has made it possible to estimate dominance genetic variance based on statistical methods, using covariance between animals predicted from genomic data. The paper aims to estimate dominance variance for some production traits in Australian sheep based on genomic data and to evaluate whether accounting for dominance effects will have an effect on the accuracy of predicting additive genetic effects. We will use data on purebred Merino sheep as well as crossbreds of Merino and other breeds to predict breeding values of purebred rams.

Materials and Methods

Phenotypes. We used 4,123 and 4,099 records for purebreds and 3,978 and 4,173 records for crossbreds for WWT and PWWT, respectively.. The animals were born between 2008 and 2011 in the Information Nucleus Research Flock (INF) of the Sheep Cooperative Research Centre (Armidale, Australia). The INF consisted of 9 flocks located across different sheep production regions in Australia. The flocks are linked to each other by using ~50% of common sires through artificial insemination. Purebred animals were Merinos and the breed composition of crossbred animals was 39.9% Merino, 24.9% Border Leicester, 22.6% Poll Dorset, 9.2% White Suffolk and 3.4% other breeds. More information on design of INF flocks is available in Van der Werf et al. (2010).

Genotypes. Animals were genotyped using the 50K Ovine SNP chip (Illumina Inc., SanDiego, CA, USA). The total number of SNP genotypes of this chip was 54,241 which decreased to 48,599 after applying quality control on genotypes. More information on quality control is available in Moghaddar et al. (2013). After quality control the missing genotypes were imputed using Beagle software (Browning, (2007)).

Statistical methods. The following linear mixed model was used for estimation of variance components.

$$y = Xb + Z_a a + Z_d d + Ww + ZQq + e$$

In this model $\boldsymbol{y}$ is the vector of phenotypes, $\boldsymbol{b}$ is the vector of fixed effects, a is the vector of random additive genetic effects, d is the vector of random dominance effects, w is the vector of random maternal effects, q is the vector of breed effects. X, Z_a, Z_d , and W are incidence matrices that relate effects to phenotypes. Q is a breed proportions matrix and $\boldsymbol{e}$ is a vector of random residual effects. a, d, w and e are normally distributed as: $a \sim N(0, \delta_a^2 G)$, $d \sim N(0, \delta_d^2 D)$, $w \sim N(0, \delta_w^2)$ and $e \sim N(0, I\delta_e^2)$, respectively and G and D are additive and dominance genomic relationship matrices**.** The fixed effects of the final model were birth type, rearing type, gender, age at measurement, and cohort of flock x birth year x management group. Breed proportion was fitted as a random effect. Analysis of the data was based on fitting additive effect (A) or additive and dominance effect (A+D). A and D are the summation of additive and dominance effect at all markers across the genome. ASReml software (Gilmour, (2009)) was used for analysis of the data. .

Additive and dominance relationships. The additive genetic relationship matrix (G) was calculated based on VanRaden (2008) as below:

$$G = ZZ' / 2\sum(p_j)(1 - p_j)$$

In this equation Z is a matrix of the size *n x m* (the number of individuals by the number of SNPs) and elements are equal to $(-2p_j)$, $(1-2p_j)$ and $(2-2p_j)$ for genotypes (A_1A_1), (A_1A_2) and (A_2A_2) of the j^{th} SNP marker genotype p_j is the frequency of the second allele (A_2) for the j^{th} SNP genotype. The dominance genomic relationship matrix (D) was calculated according to the dominance effects at each locus. This has been presented by Vitezica et al (2013), Wang et al. (2013) and Nishio & Satoh. (2014) as follows:

$$D1 = WW' / 2\sum(p_j q_j)$$

In this equation W is a matrix with elements $(-2q_i^2)$ if the genotype i of individual j is (A_1A_1), $(2p_i q_i)$ if the genotype is (A_1A_2) and $(-2p_i^2)$if the genotype is (A_2A_2). p_j and q_j are allele frequency at the j^{th} locus.

Accuracy of genomic prediction. To test the accuracy of genomic prediction the genomic breeding value (GBV) was calculated for genotyped purebred sires based on the GBLUP solution. The accuracy was evaluated as the correlation between GBV and EBV in a bivariate analysis. The EBV of those validation rams was generally based on progeny test information and calculated from a separate analysis that was ignoring phenotypes of animals used in the genomic prediction reference population.

Results and Discussion

Summary of phenotypic data. The number of phenotypes in purebred animals was 4,123 and 4,099 for WWT and PWWT, respectively, and in crossbred animals was 3,987 and 4,173 for WWT and PWWT, respectively. The phenotypic mean and standard deviation of WWT and PWWT was higher in crossbreds than in purebreds; (28.0 ±7.73) vs 24.48 ±5.25) for WWT and (45.61±7.91 vs 37.11±7.72) for PWWT, respectively.

Variance components. Table 1 shows the estimated variance for additive, dominance and residual effect. It also shows the heritability and the ratio of dominance variance to phenotypic variance for purebred and crossbred animals. The maternal effect variance estimate changed slightly between two fitting models and was lower in the A+D model. This variance is not reported in Table 1. The results show the ratio of dominance to phenotypic variance for purebred Merinos was 3.61% and 5.58% for WWT and PWWT, respectively. Those values were higher in crossbreds and were 9.21% and 17.12% for WWT and PWWT, respectively. A higher dominance variance component was expected for crossbreds and could be attributed to the higher heterozygosity and the more heterogeneous genetic structure of crossbreds compared to purebred animals. In both purebred and crossbred data the dominance variance was higher for PWWT, which is because of overall higher genetic variance in PWWT than WWT. The results also showed the residual variance was always lower when both additive and dominance effect were fitted in the model, more importantly in crossbred data. The Likelihood Ratio Test (LRT) between additive model and additive plus dominance model was highly significant in crossbred data ($p<0.01$) but was not significant in purebred data. The results also showed that the additive plus dominance model gives a reduced genetic group variance, which indicates that part of the dominance variance would otherwise be absorbed by genetic group effects.

Table 1. Variance component, narrow sense heritability and ratio of dominance variance to phenotypic variance in purebred and crossbred data.

Trait	Mod	$V_{(A)}$	$V_{(D)}$	$V_{(R)}$	h^2_A	$V_{(D)}/V_{(P)}$
WWT (p)	A	2.89 (0.48)	---	6.04 (0.35)	0.23 (0.03)	---
	A+D	2.88 (0.41)	0.40 (0.56)	5.70 (0.59)	0.23 (0.03)	0.036 (0.04)
PWWT (P)	A	7.94 (0.80)	---	9.65 (0.62)	0.36 (0.04)	---
	A+D	7.95 (0.80)	1.05 (0.96)	8.73 (1.02)	0.36 (0.04)	0.056 (0.04)
WWT (C)	A	2.47 (0.55)	---	9.82 (0.50)	0.15 (0.03)	---
	A+D	2.50 (0.54)	1.40 (0.86)	8.24 (0.78)	0.16 (0.03)	0.092 (0.05)
PWWT (C)	A	12.20 (1.27)	---	13.08 (0.85)	0.32 (0.04)	---
	A+D	11.66 (1.27)	5.94 (1.38)	8.76 (1.21)	0.32 (0.04)	0.172 (0.04)

Mod = model, WWT = weaning weight, PWWT = post weaning weight, (P) = refers to purebred animals. (C) = refers to crossbred animals. A = additive effect. D = dominance effect. P = phenotypic variance.
[1] = Numbers in parenthesis shows the standard error of estimates.

Table 2 shows the GBV accuracy for the additive and the additive plus dominance model for purebred and crossbred data. The accuracy of GBV based on an additive model or an additive plus dominance model was similar in

purebred animals. The estimation of dominance variance in purebred data was also low. For crossbred data the accuracy of GBV from fitting both additive and dominance effect increased between 0.3% and 2.1%. This coincides with higher dominance variation in crossbred data. The extra accuracy of including dominance effects in the model was observed in all cases, but was low and statistically not significant ($p<0.05$). One explanation is that the additive genetic variance will absorb a lot of the variation due to dominance effects. This has been pointed out by Falconer (1981) and Hill et al. (2008). However, the amount of non-additive genetic variance could be different across different traits and some studies have reported substantial non-additive genetic variance in some polygenic traits (e.g. Gengler et al. (1997); Palucci et al. (2007)). Furthermore, dominance effects can also mask additive genetic differences, especially when they are expressed in crossbreds, and this might explain that accounting for them improves the accuracy of genomic prediction of additive genetic effects. The presented results here are still associated with considerable standard error, and more analysis on more traits and larger datasets is required to confirm these findings.

Table 2. Accuracy of GBV for additive and additive plus dominance model based on purebreds or crossbreds reference population.

Trait	Mod	P.B Ref.Pop	C.B Ref.Pop			
		Mer	Mer	BL	PD	WS
WWT	A	0.462	0.357	0.453	0.264	0.234
	A+D	0.463	0.376	0.461	0.271	0.244
PWWT	A	0.550	0.462	0.435	0.273	0.183
	A+D	0.551	0.483	0.438	0.286	0.192

Mod = model, WWT = weaning weight, PWWT = post weaning weight, (A) additive model. (A+D) additive plus dominance model. Mer = Merino, BL = Border Leicester, PD = Poll Dorset, WS = White Suffolk, P.B Ref-Pop = purebred reference population C.B Ref.Pop = crossbred reference population.

The estimated dominance variance is affected by marker allele frequencies (Falconer, (1981)). It is expected that the allele frequency of QTL are more extreme compared to SNP marker allele frequency. Therefore, analysis with denser marker data, or based on genome sequence data could give better estimates of dominance variance and its effect on genomic prediction accuracy.

Conclusion

This study estimated dominance variation of weaning and post weaning traits based on genomic information in real purebred and crossbred sheep data. The dominance effect was significantly higher in crossbred data compared to purebred data. Fitting an additive plus dominance effect model provides similar or higher accuracy of genomic breeding value, particularly for traits with higher dominance variation, which is more important in crossbreeding or mate selection programs.

Literature Cited

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

Falconer, D. S. (1981) Introduction to quantitative genetics. Longman.
Genetics Selection Evolution 39: 181–193.

Gengler, N, VanVleck, L.D., MacNeil, M. D., et al,(1997) *J. Anim. Sci.* 76: 2515–2520.

Gilmour, A.R., Gogel, B.G., Cullis, B.R.,Thompson, R., (2009): ASReml User Guide Rrelease 3.0. Hemel Hempstead:VSN International Ltd

Hill, W. G., Goddard, M.E.,Visscher, P. M., (2008),. PLoS *Genetics 4(2)*: e1000008.
J. Anim Sci 75: 2885–2891.

Misztal, I., Varona, L, Culbertson, L. et al (1998) *Biotech. Agro.Soc.Env.* 2: 227-233.

Moghaddar, N,. Swan. A. A., Van der Werf. J. H. J. (2013) *Anim. Prod. Sci.* [http:/dx.doi.org /10.1071/AN13129]

Moghaddar, N. Van der Werf, J.H.J (2014). *Genet. Sel. Evol.* (submitted)

Nishio,. M. Satoh, M., (2014) *Plos-one* DOI: 10.1371/journal.pone.0085792

Palucci, V., Schaeffer L.R., Miglior, F., Osborne, V. (2007) *Genet. Sel. Evol* .39: 181–193.

Van der Werf, J. H., Kinghorn, B.P., Banks, R.G.(2010). *Anim. Prod. Sci*, 50:998-1003

VanRaden, P. M., (2008). J. Dairy Sci. 91: 4414-4423.

Vitezica, Z.G., Varona, L., Legarra, A. (2013) Genetics:10.1534/genetics.113.155176

Wang, C. et al (2013) . J Dairy Sci 96 (E-Suppl (1)) 620