

**THE RELATIVE IMPORTANCE OF INFORMATION ON UNRELATED INDIVIDUALS
ON THE PREDICTION OF GENOMIC BREEDING VALUES**

S.A. Clark^{1,2}, J.M. Hickey¹ and J.H.J. van der Werf^{1,2}

¹University of New England, Armidale, NSW 2351

²CRC for Sheep Industry Innovation, University of New England, Armidale, NSW 2351

SUMMARY

The theory of genomic selection is based on the prediction of the effects of genetic markers in linkage disequilibrium (LD) with quantitative trait loci (QTL). However, there is increasing evidence that genomic selection also relies on relationships between individuals or the patterns of LD associated with these relationships to accurately predict genetic value. This study aimed to examine the relative importance of information on essentially unrelated individuals on the estimation of breeding value when using gBLUP and BLUP.

Analysis was undertaken using a simulated population of 2000 animals. Two reference populations were formed from 1750 animals and the accuracy of prediction was assessed for the remaining 250 animals that formed the test population. Two test populations were constructed such that one included 10 families that had no family members in the reference population and the other included 5 half siblings from 50 families. The gBLUP method more accurately predicted breeding value than BLUP in both test populations. The highest accuracy was achieved when gBLUP was used to predict the breeding value of closely related animals. However, gBLUP was still able to predict breeding value accurately even when animals were distantly related.

INTRODUCTION

Genomic selection (GS) is a method to predict breeding values in livestock; however the mechanism by which it predicts is not completely clear. Initially it was thought that GS predicted effects of genetic markers in linkage disequilibrium (LD) with quantitative trait loci (QTL) (Meuwissen *et al.* 2001). However, there is increasing evidence that genomic selection also relies on relationships between individuals to accurately predict genetic value because predictions are more accurate when they occur between highly related populations (Habier *et al.* 2010).

The LD/QTL paradigm would suggest that accurate predictions of breeding value would persist for several generations into the future allowing for a reduced number of phenotypic measurements in each generation (Muir, 2007). In contrast, if the predictive ability of GS is based on close pedigree relationships between animals, genomic predictions may remain accurate for only one or two generations and continuous measurement of phenotypes of individuals that are related to selection candidates would be needed. The question arises; does an animal that has its breeding value predicted from genomic information require close relatives in a reference population?

There are various methods used for predicting breeding values from genomic data. These range from; Bayes B which allows each locus to explain different amounts of variation, with only a small number of loci having an effect and many loci are assumed to have no effect (Meuwissen *et al.* 2001) to gBLUP which assumes equal variance across all loci (Habier *et al.* 2007). However, empirical evidence across livestock populations has shown that in many cases these methods obtain very similar accuracies of the estimated breeding value (Moser *et al.* 2010).

Genomics

Many studies have been published regarding the use of the gBLUP method to predict breeding value (VanRaden *et al.* 2009, Hayes 2009, Moser *et al.* 2010). gBLUP uses genomic information to form a genomic relationship matrix (GRM) that defines the additive genetic covariance between animals (Nejati-Javaremi *et al.* 1997). The GRM then replaces the numerator relationship matrix (NRM) in the traditional BLUP equations, which is based on pedigree relationships. The GRM is expected to give a more accurate estimate of covariance. However, it is relatively unknown how much accuracy is gained from improved measures of covariance among known relatives and how much is gained from information on distant ‘relatives’ previously ignored via the pedigree method. The aim of this study was to observe the relative importance of information on essentially unrelated individuals on the estimation of genomic breeding values.

METHODS

Base Genotype Simulation: Genotype simulations were conducted using the Markovian Coalescence Simulator (MaCS) (Chen *et al.* 2009) to simulate the base haplotypes of single nucleotide polymorphism (SNP) sequence data, which were then allocated to a simulated population structure. Phenotypes were simulated under a traditional QTL model with 1000 QTL as defined in Clark *et al.* (2010). Each SNP in the sequence had a 3% chance of being used as a marker and a 0.05% chance of being used as a QTL. The population was simulated for 10 generations and each generation contained 4000 animals, half male and half female. Eighty males were randomly selected in each generation and randomly mated to all females which each had two offspring per generation.

Analysis of data. A random selection of 60,000 SNP markers was used in the genomic evaluation. Genomic evaluation was undertaken using the gBLUP method using a genomic relationship matrix to define covariance between the animals in the population. The GRM was formed as defined in VanRaden (2008). Similarly, traditional best linear unbiased prediction (BLUP) was performed using a deep (BLUP-D), 10 generation pedigree and a shallow, single generation pedigree (BLUP-S). Each scenario was replicated 8 times.

The empirical accuracy ($r_{(cor)}$) of the breeding values estimated in the test set was defined as the correlation between the true and estimated breeding value. The accuracy was also derived for each individual as: $r_{(PEV)} = \sqrt{(1 - (PEV/G_{ii}V_a))}$ where; PEV is the prediction error variance estimated using elements from the mixed model equations. G_{ii} is the diagonal of the GRM for animal i and is substituted for A_{ii} in traditional BLUP, V_a is the additive genetic variance for the population. Furthermore, $PEV = C_{ii}V_e$ where; C_{ii} is the diagonal of the coefficient matrix for animal i and V_e is the residual variance.

All analyses were undertaken on 2000 animals from the final generation. A reference population was formed from 1750 animals and the accuracy of prediction was assessed for the remaining 250 animals that formed the test population. Two test populations were constructed such that each population had 250 animals. Test population 1, included 10 families that had no direct family members in the reference population. And test population 2, included 5 half siblings from 50 families such that each animal had 20 half siblings in the reference population.

RESULTS

Breeding values that were estimated using gBLUP achieved the highest accuracy. When animals in the test population and reference populations were closely related the highest accuracy was reached (Table 1). When the two populations were not closely related, accuracies were generally lower but the reduction was much smaller for gBLUP, which gave a much higher accuracy than BLUP, in fact a similar accuracy to that achieved by BLUP-D (deep pedigree) with closely related animals.

Table 1. The empirical accuracy ($r(\text{cor})$) of the estimated breeding value when animals in the test set were closely and distantly related to animals in the reference population.

Method	Distant*		Close	
BLUP-S	0.00	(0.000)	0.39	(0.021)
BLUP-D	0.21	(0.031)	0.42	(0.019)
gBLUP	0.41	(0.034)	0.57	(0.014)

BLUP-D gave low accuracy when there were no first degree relationships between animals in the reference and test populations. However when no pedigree was used, BLUP-S (shallow pedigree) was unable to estimate breeding value. The deep pedigree used in BLUP-D enabled the estimation of a proportion of covariance between the test and reference populations based on information from distant ancestors. In contrast when relatives were present in both populations BLUP was able to predicted breeding value quite accurately regardless of the length of the pedigree.

The estimation of accuracy, $r_{(PEV)}$, when averaged over the test population was similar to the empirical accuracy of the group $r(\text{cor})$. However, for gBLUP the theoretical accuracy under-estimated realized accuracy when family information was used (Figure 1).

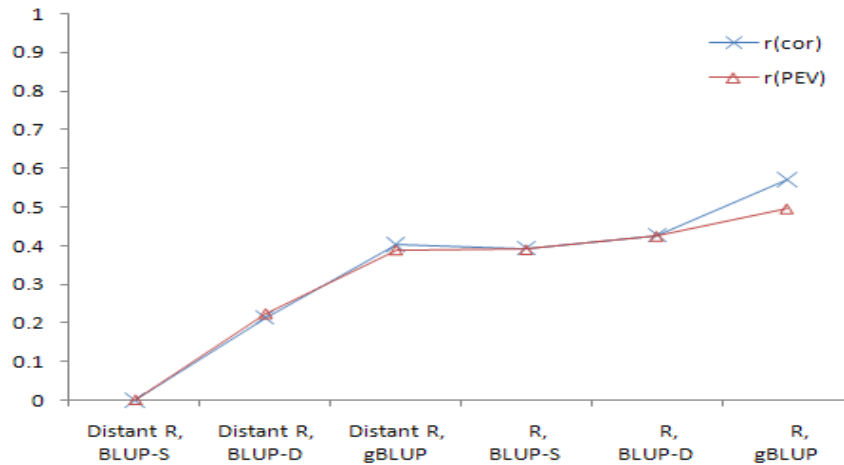


Figure 1. The estimation of accuracy based on the PEV from the coefficient matrix ($r_{(PEV)}$) and the correlation between estimated and true breeding value ($r_{(COR)}$) for all scenarios (R= Relatives).

* Standard error of means

The results show that when there is a distant relationship between the animals in the test and reference populations, gBLUP is still able to accurately predict breeding value. When no other information is available, the all of the information gathered from distantly related animals contributes to the accuracy of prediction. However when relatives are included in the reference population, it is likely that the importance of information on distantly related animals is reduced. Selection index theory shows that when information on closely related animals is available, more weight is placed on this information and therefore information from distantly related animals becomes less important. Although the importance of information from distant relatives is reduced this extra information still enables gBLUP to achieve a higher accuracy than BLUP-D. Furthermore, when gBLUP is compared to BLUP-S, which only uses information on close relatives, the extra information on the distantly related animals contributes to an 18% increase in accuracy.

The relative weights placed on information from distantly and closely related animals may have important implications when assessing the makeup of reference populations and on the duration of the response from GS in genomic breeding schemes. The inclusion of information on relatives will improve the accuracy of the predicted breeding value. However, gBLUP is still able to use information on distantly related animals to give a relatively accurate prediction of breeding value.

CONCLUSION

The relationships between animals affect the accuracy of predicting breeding value using gBLUP. When there is a close relationship between the animals in the reference and test populations, gBLUP can estimate breeding values with a high accuracy. However, even when there is only a distant relationship between the animals in test and reference populations, gBLUP is still able to give an accurate estimate of breeding value.

REFERENCES

- VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F. and Schenkel F. (2009) *J Dairy Sci*, **92**:16.
- Chen G.K., Marjoram P. and Wall J.D. (2009) *Genome Res.* **19**: 136.
- Villa-Angulo R, Matukumalli L.K, Gill C.A, Choi J, Van Tassell C.P, and Grefenstette J.J. (2009) *BMC Genetics*, **10**:19
- Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) *Genetics*, **157**:1819.
- Muir W.M. (2007) *J Anim Breed Genet*, **124**: 342.
- Nejati-Javaremi, A., Smith C. and Gibson J. P. (1997) *J. Anim. Sci.* **75**: 1738.
- Clark S.A., Hickey J.M. and van der Werf J.H.J. (2010) *Proc 9th WCGALP*. 0944:1.
- Moser G., Tier B., Crump R.E., Khatkar M.S. and Raadsma H.W. (2009) *Genet. Sel. Evol*, **41**:56.
- Hayes B.J., Bowman P.J., Chamberlain A.C. and Goddard M.E. (2009) *J Dairy Sci*, **92**:433.
- VanRaden P.M. (2008) *J Dairy Sci*, 33.
- Habier D., Tetens J., Seefried F.R., Lichtner P. and Thaller G. (2010) *Genet. Sel. Evol.* **42**:5.