

IMPLICATIONS OF GENETIC ARCHITECTURE ON THE EFFICACY OF GENOMIC SELECTION

F.S. Hely¹, S.A. Clark² and P.R. Amer¹

¹AbacusBio Limited, P O Box 5585, Dunedin 9058, New Zealand

²School of Environmental and Rural Science, University of New England, Armidale, NSW, 2351, Australia

SUMMARY

A simulation model is described which has been constructed to address the issue of how true underlying genomic architecture might impact on the efficacy of genomic selection. A current specific focus of the model is on how epistatic genetic architectures might impact on the added value expected from increasing the density of SNP markers. Results to date suggest that genomic selection has greater superiority over BLUP genetic prediction under the additive genetic architecture simulated relative to an epistatic architecture with similar heritability. While we expect marker density to improve accuracy under GBLUP with some additive genetic architectures, our simulation results suggest that this may not happen with comparable (in terms of narrow sense trait heritability) genetic architectures with epistatic gene action contributing to both additive and non-additive genetic variance.

INTRODUCTION

The underlying genetic architecture of economically important traits in sheep remains unclear. There is a reasonable body of biological evidence (Gianola and de los Campos 2008) that suggests interacting genetic loci (i.e. epistatic loci) are a significant source of genetic variation. It is yet to be determined how single step genomic selection will perform when epistatic effects among loci contribute significantly to underlying additive genetic variation. The genomic best linear unbiased prediction (GBLUP) method of genomic selection assumes each SNP marker has an equal effect on trait variance and uses information from the genomic relationships between candidates to estimate the merit of genotyped candidates as opposed to alternative Bayes methods which use the effects of minor and major genes weighted differently. In this study, simulation work was undertaken to model the application of single step genomic selection methodology to the New Zealand sheep industry using a combination of low and high density SNP panels. A set of QTL were simulated, and the accuracy of prediction using both conventional BLUP genetic evaluation and the single step GBLUP genetic evaluation was compared with and without epistatic genetic effects simulated for a single trait in a population resembling a major NZ dual purpose sheep breed.

METHODOLOGY

Population and SNP data were simulated using the QMSim software developed by Sargolzaei and Schenkel (2009). The parameters used in the simulations are shown in Table 1. These parameters were chosen to try and generate a population with similar characteristics to the major New Zealand dual purpose sheep breeds. The QMSim software uses a two stage method for simulating a population; a historical phase and a recent population phase. The historical phase uses random mating over a large number of generations to create linkage disequilibrium and drift in a base population. The recent population phase is used to create the desired population structure for analysis, no mutation occurs and the allele effects are fixed at the end of the historical phase.

Table 1: Parameter estimates for the population simulated using QMSim

| Parameter | Value |
|---|-----------------------|
| Effective population size for the historical phase | 4000 |
| Number of females per male in the historical phase | 20 |
| Number of generations for the recent population | 60 |
| Number of females per male in the recent population | 50 |
| Litter size in the recent population | 50% single, 50% twins |
| Proportion of male progeny in the recent population | 0.5 |
| Replacement ratio for sires/dams | 1.5 yrs/3 yrs |
| Number of chromosomes | 26 |
| Marker and QTL mutation rates | 2.5×10^{-5} |

Once the QMSim data were generated, epistatic and purely additive true breeding values were simulated for all individuals with marker data available. The additive true breeding values (TBV_{add}) were calculated using the sum of the allele effects provided by QMSim for 100 QTL segregating at the end of the historical phase. These QTL had additive effects which were sampled from a normal distribution. The epistatic true breeding value (TBV_{epi}) was calculated in a similar way. For n pairs of loci with epistatic effects simulated between them n 9x9 matrices of epistatic effects for all possible combinations of genotypes were simulated. For a given pair of loci A and B each with two alleles (a and A , b and B respectively) a matrix was created as below:

$$\begin{array}{cccc}
 & bb & bB & BB \\
 aa & e_{ab} & 0 & e_{aB} \\
 aA & 0 & 0 & 0 \\
 AA & e_{Ab} & 0 & e_{BB}
 \end{array}$$

Thus, if an individual had the combination of genotypes aa and BB then $TBV_{epi} = TBV_{epi} + e_{aB}$. The epistatic effects e were drawn from a normal distribution. In order to compare genomic breeding values based on additive versus epistatic true breeding values, it was necessary to scale the variance of the true breeding values so that the additive genetic variance estimated by ASReml (Gilmour *et al.* 1999) was the same for both the epistatic and additive genetic models. i.e.

$$TBV^* = TBV \times \frac{\sqrt{h^2}}{\sigma_{TBV_N}}$$

where TBV^* is the rescaled TBV , h^2 is the desired trait heritability, σ_{TBV_N} is the additive genetic standard deviation (narrow sense) estimated by ASReml. Phenotypes were then simulated as

$$PHEN = TBV^* + (1 - h^2) \times \frac{\sigma_{TBV_B}^2}{\sigma_{TBV_N}^2} \times \delta$$

where σ_{TBV_B} is the standard deviation of the original TBV s in the broad genetic sense and δ is a random normal deviate with mean of 0 and standard deviation of 1. In this way, the two different

architectures are constructed in such a way that they would appear to be identical when undertaking variance component estimation using conventional quantitative genetic analysis.

A genomic best linear unbiased prediction GBLUP evaluation was run on the phenotypic values for both the additive and epistatic traits using the BLUPF90 family of programs (Miszta *et al.* 2002) with a SNP marker file. A traditional BLUP evaluation was also run using ASReml and the estimated breeding values from both evaluations were combined with the TBV and phenotypic data. Accuracies of genomic predictions were computed as the correlation between the additive and epistatic TBVs and their corresponding genomic estimated breeding values (GEBVs).

SNP panel densities from 10,000 to 100,000 were simulated, with the accuracies, measured as the correlation of the TBVs with the GEBVs and BLUP EBVs, for the different panel densities compared. The TBVs were scaled to give an additive genetic variance of 0.3. The number of QTL used to generate the additive and epistatic TBV remained constant at 100 for all scenarios. From QMSim, the marker data were retained for individuals generated in generations 57 to 60. For a training and validation trial, the individuals born in generation 57 had phenotypic data and all other individuals had a missing phenotype. Correlations between estimated breeding values and true breeding values are reported for animals from generation 60.

For all scenarios 20 replicates were run, where replication was performed by using the same base population markers and pedigree from QMSim for the 20 replicates, but with a new true genetic values for each replicate. Within each replicate, the GBLUP, Bayes Lasso and pedigree BLUP methods are applied to the exact same trait data with the same model.

RESULTS AND DISCUSSION

The accuracy of genomic selection (as indicated by correlations between predicted breeding values and true breeding value) exceeded the accuracy of BLUP genetic predictions for animals in the validation population which did not have their own phenotypic records (Table 2). BLUP genetic predictions appeared slightly more accurate under the additive model than under the epistatic model although the difference was not statistically significant. In contrast, genomic prediction was much more accurate under the additive model than under the epistatic model. Increasing the SNP density from 5k to 100k did not have any meaningful impact on the results with these genomic architectures and population structures.

Table 2: Correlations and the standard errors between true and estimated breeding values using GBLUP (TBV-GEBV) and traditional BLUP (TBV-EBV) for additive and epistatic traits, along with the heritability as estimated by ASReml with the standard error (simulated heritability was 0.3 for all scenarios).

| Panel Size | Additive | | | Epistatic | | |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | TBV- GEBV | TBV-EBV | herit | TGV- GEBV | TGV-EBV | herit |
| 5k | 0.65 (0.004) | 0.36 (0.002) | 0.38 (0.015) | 0.6 (0.009) | 0.28 (0.013) | 0.3 (0.012) |
| 10k | 0.69 (0.005) | 0.43 (0.004) | 0.34 (0.012) | 0.60 (0.01) | 0.26 (0.012) | 0.33 (0.012) |
| 20k | 0.75 (0.003) | 0.45 (0.003) | 0.36 (0.007) | 0.62 (0.01) | 0.31 (0.01) | 0.3 (0.009) |
| 50k | 0.74 (0.002) | 0.39 (0.005) | 0.32 (0.007) | 0.64 (0.009) | 0.30 (0.014) | 0.3 (0.012) |
| 100k | 0.75 (0.004) | 0.49 (0.003) | 0.30 (0.011) | 0.60 (0.01) | 0.33 (0.012) | 0.28 (0.012) |

We hypothesise that with further exploration of population structures and genomic architectures, we will find situations where increasing marker density will increase the accuracy of

genomic predictions under the additive genetic architecture, but they will be less beneficial under the epistatic genetic architecture. This is because similarity among relatives due to sharing equivalent epistatic gene combinations breaks down much more quickly over successive meioses than similarity due to inheritance of similar additive genetic effects. It is acknowledged that some patterns within the results appear inconsistent with the relative small sizes of standard errors. We believe that this may be due to replication being undertaken with the same set of SNPs.

CONCLUSION

If our hypothesis is confirmed through further work, then new approaches other than GBLUP, Bayes predictions, and single step genetic evaluation may be required to capture the full benefits from increased marker density when traits whose observed narrow sense heritability is driven by epistatic effects. Alternatively, the failure of Bayes methods, and increased marker density to meaningfully improve the accuracy of genomic selection in many practical situations tested to date, could be further evidence that epistasis is an important contributor to observed heritability in livestock populations. The alternative theory of many genes with very small effects has led to considerable, but so far fruitless, efforts to use increasingly dense marker chips to improve genomic selection both within and across breeds beyond what can be achieved with moderate density chips (e.g. 50k).

ACKNOWLEDGEMENTS

We gratefully acknowledge funding for this work by Ovita Limited.

REFERENCES

- Gianola, D. and de los Campos, G. (2008) *Genetics Research* **90**: 525.
Gilmour A.R., Cullis B.R., Welham S.J. and Thompson R. (1999) *NSW Agriculture Bulletin* No. **3**.
Misztal I., Tsuruta S., Strabel T., Auvray, B., Druet T. and Lee D. (2002) *Proc.7th WCGALP* 28.
Sargolzaei, M. and Schenkel F.S. (2009) *Bioinformatics* **25**: 680.