Proc. Assoc. Advmt. Breed. Genet. 21: 137-140

USING GENOMIC INFORMATION TO ESTIMATE GENOTYPE BY ENVIRONMENT INTERACTIONS

S.A. Clark[1,2] , S.H. Lee[1,2] and J.H.J. van der Werf[1,2]

[1] School of Environmental and Rural Science, University of New England, Armidale, NSW 2351
[2] CRC for Sheep Industry Innovation, University of New England, Armidale, NSW 2351

SUMMARY

Genotype by environment (GxE) interaction can reduce genetic gain because there is often insufficient information for accurate selection in each environment. Traditionally the estimation of GxE effects has been based on the performance of half siblings across environments. This limits the estimation of GxE to specifically designed datasets with close relatives where all realized relationships may not be utilized. Genomic information can also be used to link animals and presents an opportunity to compare genotypes across different environments using realized relationship information. This study examines the use of genomic information to estimate GxE interaction. The genetic correlation between animal phenotypic performance in two different environments was estimated using pedigree or genomic information. A higher genetic correlation between environments was observed when using genomic information (0.9) than when using pedigree information (0.71). This study suggests that genomic information may be a useful alternative to pedigree information in understanding GxE in livestock populations.

INTRODUCTION

In livestock production, animals are recorded and selected in a wide range of environments. While for most economically important traits there is little evidence for genotype by environment interaction (GxE), for some traits, animals or genotypes may perform differently in each environment (i.e. across different geographic locations or from one year to another). This can involve a change in the differences between alternative genotypes (often referred to as scale effects) and it can also relate to a change in the ranking of genotypes across environments.

Traditionally, GxE interactions can be estimated by measuring relatives across environments. Genotype by environment interactions can be estimated using mixed model analyses treating performance across environments as two different traits (Falconer 1952) and estimating the genetic correlation between performances across each environment. Past studies examining GxE have been limited to experimental designs that primarily focus on the use of common sires across various environments. The advent of genome-based technologies allows for the possibility of changing the way GxE interactions may be estimated.

Genome-wide association studies and genomic prediction have become common place for the prediction of disease risk in human populations and for predicting genetic merit in livestock (Goddard 2012). Genome-wide association and genomic prediction rely on a group of individuals with both genotypic and phenotypic information to enable the prediction of marker effects (directly or indirectly). Often these phenotypes come for a wide range of environments and the genomic information can be used to define the covariance between relatives (in the form of a genomic relationship matrix (GRM)) (VanRaden 2008). Genomic information presents an opportunity to enable a more diverse range of animals, not just close pedigree relatives, to contribute to estimating a genetic correlation between environments. It also presents the opportunity to observe whether specific genomic regions are more important than others for performance in varying environments.

The aim of this study was to use genomic information to estimate GxE and to examine the impact of such information when compared to pedigree based estimates.

METHODS

The data used in this study consisted of phenotypic and genotypic records from Merino animals in the Australian Sheep Cooperative Research Centre (CRC) information nucleus flock (INF). The INF is a specifically designed dataset that includes animals that have been recorded in eight environments across Australia. This dataset consisted of a dataset of phenotypic and genotypic records from 4433 Merino animals for the Post Weaning Weight (PWWT) trait. This dataset was further broken down such that phenotypic data from 1807 animals from 227 sires measured across two environments were extracted for the analysis. Location 1 (E1), Armidale (NSW) is a temperate environment with a primarily summer-dominant rainfall (n=921) and Location 2 (E2) Katanning (WA) is located in a winter-dominate rainfall zone (n=886).

All animals in each dataset were genotyped using the Illumina 50K ovine SNP chip. All SNP in this dataset underwent a number of genotyping quality control measures (see Daetwyler et al. (2010) and 48 599 markers remained following the quality control.

Genotype by environment interaction was estimated using both pedigree and genomic information. Phenotypic performance in the two environments was modelled as two separate traits. A bivariate animal model was fitted in ASReml (Gilmore 2009) and the genetic correlation between performance across environments was estimated using either a genomic or pedigree based relationship matrix. The following fixed effects were fitted in the analysis of PWWT: Sex, birth type, rearing type, age of dam, contemporary group (birth year • site • management group) and age-at-trait recording. We assumed the following model;

$$y_i = X_i b_i + Z_i a_i + Q_i s + e_i \quad (1)$$

where $\mathbf{y}_i$ is a vector of phenotypes for environment *i*, $\mathbf{X_i}$ is a design matrix relating the fixed effects (as described above) to each animal for environment *i*, $\mathbf{b}_i$ is a vector of fixed effects, $\mathbf{Z}_i$ is a design matrix allocating records to breeding values, **a** is a vector additive genetic effects for animals, $\mathbf{Q}_i$ is a matrix relating animals to genetic groups and **s** is a vector of genetic group effects and $\mathbf{e_i}$ is a 2x2 diagonal matrix of random normal deviates $\mathbf{I}\sigma^2_{ei}$. Furthermore V(a) = $\begin{bmatrix} \mathrm{A}\sigma^2_{a1} & \mathrm{A}\sigma_a \\ \mathrm{A}\sigma_a & \mathrm{A}\sigma^2_{a2} \end{bmatrix}$ where σ^2_{ai} if the genetic variance for environment *i* and σ_a is the covariance between environments and A is the numerator relationship matrix. In the genomic analysis, the genomic relationship matrix (GRM) replaced the A such that V(g) = $\begin{bmatrix} \mathrm{G}\sigma^2_{g1} & \mathrm{G}\sigma_g \\ \mathrm{G}\sigma_g & \mathrm{G}\sigma^2_{g2} \end{bmatrix}$ (VanRaden 2008).

Marker effects for each environment were also estimated using single marker regression using the R package lm. The model fitted was

$$y = Xb + Q + SNP_j + e_i \quad (2)$$

Again **y** is a vector of phenotypes, **X** is a design matrix relating the fixed effects (as described above) to each animal. Genetic groups (**Q**) were fitted as fixed effects. Each SNP was individually fitted until all markers had been tested. In this analysis three groups of phenotypic data were used to estimate the marker effects; the complete INF dataset of Merino animals with PWWT records (n=4433) across all eight environments, records from E1 (n=921) and records from E2 (n=886). The 500 most significant markers from E1 and E2 were then used to estimate a correlation across environments.

RESULTS

A moderate genetic correlation between environments was estimated for PWWT in Merino sheep using pedigree information (Table 1). By contrast, when the GRM was used to define the

covariance between individuals the genetic correlation between performances across environment was higher. Similar variance components were estimated for E2 when using either genomic or pedigree information. However, large differences between the variance component estimates were observed for E1 which contributes to the variable genetic correlation estimates. There was a high standard error surrounding each genetic correlation such that each correlation was not significantly different however it is interesting to observe such large dissimilarities between the estimates.

Table 1. Genetic variance (V_a), Phenotypic variance (V_p), heritability (h^2) and genetic correlation (r_g) of performance across alternative environments using either pedigree or genomic information (S.E).

	Pedigree				Genomic			
	E1	**E2**	**cov**	**r_g**	**E1**	**E2**	**cov**	**r_g**
V_a	10.35	13.00	8.32	0.71 (0.18)	6.79	12.15	8.192	0.90 (0.15)
V_p	15.55	23.04			15.19	22.77		
h^2	0.66 (0.11)	0.56 (0.12)			0.44 (0.074)	0.53 (0.085)		
LogL	-3458.36				-3434.29			

The marker effects were different across environments (Figure 1). When information from eight environments was used to estimate marker effects a large significant peak was observed on Chromosome 6. This location is consistent with that reported by Al-Mamun et al (2014). The strength of this peak reduced when the dataset was limited to either location 1 or 2 information (Figure 1b and 1c). Figure 1d shows the relationship between the effects of the most significant markers from each environment (E1 and E2).

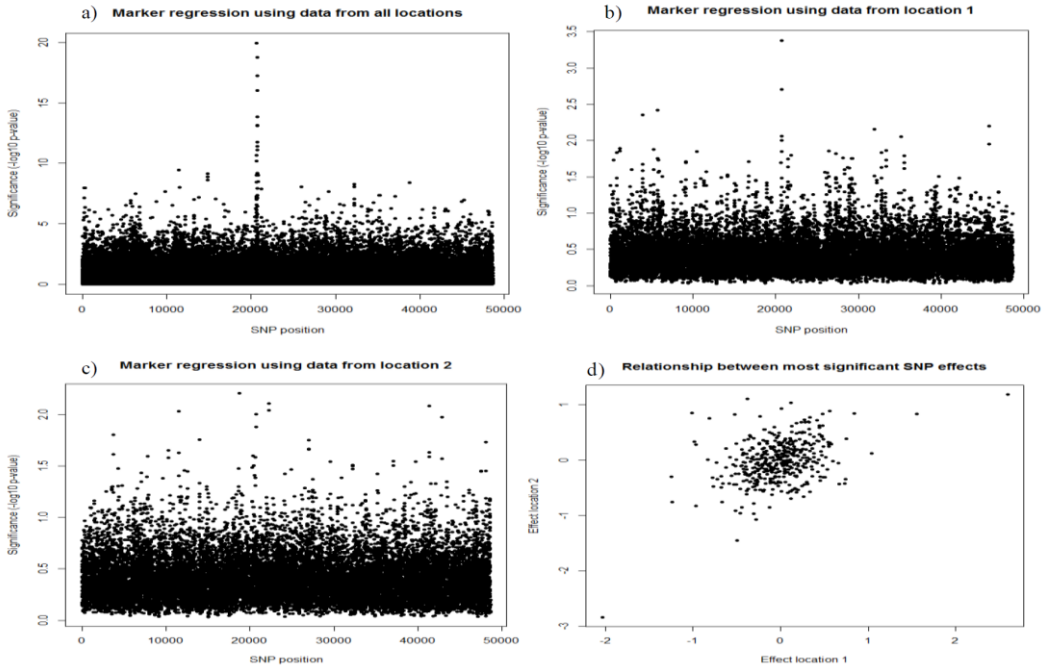

Figure 1 Manhattan plot of marker significance using data a) from all eight INF environments b) from location 1 c) from location 2. d) The relationship between the 500 most significant SNP from b & c.

The correlation between the SNP effects of significant markers estimated from each environment was lower than that estimated from the bivariate analysis (0.39). Estimating marker effects from individual environments is somewhat problematic given the reduction in the number of records available to detect marker effects, which can only be improved by increasing the amount of data used to estimate such effects. Combining data from many environments allowed for greater statistical power to be achieved and for a significant region to be observed. If a significant GxE interaction was to exist, marker effects may also be affected by this interaction and therefore may not result in consistent predictions across environments (if data was separated into specific environments). Furthermore, the estimated correlation between significant effects may not be a true reflection of the actual genetic correlation across environments due to the high degree of similarity between significant markers (i.e. many markers are in fact tracing the same genomic region). There would have also been a large amount of Linkage disequilibrium between markers due to the structure of the data. This could be corrected for by fitting all markers within the model (i.e. RR BLUP) but given the equivalence between gBLUP and RRBLUP (Habier et al 2013) the current gBLUP analysis would have resulted in a better estimate of the genetic correlation between environments.

The reasons for the disparity between pedigree and genomic estimates are not completely clear. The Log likelihood from each analysis suggests that using genomic information was in fact a better model, significantly increasing the likelihood of the data. This increase, however, may have been due to a number of factors. The first explanation is that the GRM may have better parameterized the relationship between the commercial dams that were used to create this dataset and better corrected for the genotypic effects across environments than what was captured by pedigree. A second explanation is that the GRM may have also included some genetic group information that was not available to the pedigree based matrix and could not be separated from the GRM. This would imply that the genomic estimate may be overestimating the genetic correlation across environments.

CONCLUSION

Genotype by environment interactions can be estimated using either pedigree or genomic information. Genomic information allows for the comparison of all animals across environments, not just animals from sire families or that have close pedigree links. Estimates of GxE may be different when comparing pedigree or genomic relationships and careful consideration needs to be made when interpreting such differences.

REFERENCES

Falconer, D.S. (1952). *Amer. Nat.* **86**:293–298

Gilmour, A.R., Gogel, B.J., Cullis, B.R. *et al*. (2002). ASReml User Guide Release 3.0 VSN International.

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

Daetwyler, H.D., Hickey, J.M., Henshall, J.M., Dominik, S., Gredler, B. *et al*. (2010). *Anim. Prod. Sci.* **50**: 1004.

Al-Mamun, H.A., Clark, S.A., Kwan, P., Gondro, C. (2014) Proc. 10th WCGALP.

Goddard, M.E. (2012) *Anim. Prod. Sci.* **52**: 73.

Habier D., Fernando R.L., Garrick D.J., (2013) *Genetics* **194**(3): 597.