Proc. Assoc. Advmt. Anim. Breed. Genet: 24: 135-138

MULTIVARIATE ANALYSES USING TWO GENOMIC RELATIONSHIP MATRICES TO WEIGHT PREDICTIVE SNP MARKERS

P.M. Gurman[1], L. Li[1], A.A. Swan[1], N. Moghaddar[2] and J.H.J. van der Werf[2]

[1] Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia
[2] School of Environmental & Rural Science, University of New England, Armidale, NSW, 2351 Australia

SUMMARY

The Neogen GGP Ovine 50k chip contains approximately 5000 predictive Single-nucleotide polymorphisms (SNPs) that were identified by the Sheep CRC based on their relationship with carcase traits from genome wide association studies. These SNPs have been used in routine MERINOSELECT and LAMBPLAN analyses, equally-weighted with all other SNPs in a single genomic relationship matrix (GRM). This study aimed to examine the impact of fitting all SNPs in one GRM or fitting two GRMs, one with selected predictive SNPs and one with random SNPs, in conjunction with a numerator relationship matrix. Phenotypes on terminal sire breed cross resource flock animals recorded for five carcase and eating quality traits were used for bivariate variance component estimation. Variance components estimates were obtained for models containing only a numerator relationship matrix (NRM), NRM plus a GRM containing only non-selected SNPs, an NRM plus two GRMs containing non-selected and selected SNPs and an NRM plus one GRM containing all SNPs. Log-likelihoods were significantly higher in the models containing two GRMs for all trait pairs. Slightly higher average heritabilities were estimated from the model where the GRM contained all SNPs, except for intramuscular fat and shear force, where the GRM without the predictive SNPs resulted in higher heritabilities. The proportion of genetic variance explained by the genomic relationship matrices (λ) was estimated to be between 0.59 and 0.86. In terms of the genetic correlations between traits, for many trait-pairs the correlations were similar between the random effects fitted, but for two trait-pairs large differences were observed between the genetic correlations.

INTRODUCTION

Routine genetic evaluations for Australian terminal sire, maternal and Merino sheep have utilised single-step genomic BLUP (SS-GBLUP) since 2017 (Brown *et al.* 2018). For the genomic relationship matrix used in these analyses, the SNPs used were based on a set that passed quality control from the ISAG 50k sheep panel. In 2019, a new genomic panel for sheep was introduced (GeneSeek Genomic Profiler Ovine 50k, Neogen) which included approximately 5000 additional predictive SNPs that have been significantly associated with specific growth, carcase and eating quality traits in sheep (Moghaddar *et al.* 2019). The union of all SNPs on all genomic panels was chosen (including the predictive SNPs), with imputation of missing SNPs on each panel, followed by imputing all panels to the union set, resulting in 60,410 SNPs used in SS-GBLUP.

The methods commonly used for constructing the genomic relationship matrix (GRM) for GBLUP (VanRaden 2008; Yang *et al.* 2010) assumes that all SNPs have equal weighting. While equal weighting on SNPs is reasonable for random SNPs, it may be appropriate to treat selected SNPs that are associated with specific traits differently. The GRM used in SS-GBLUP is blended with the NRM for these animals based on the parameter λ, with the currently used value in Australian sheep evaluations set to $\lambda = 0.5$ resulting in the weighted GRM being the mean of the raw GRM

* A joint venture of NSW Department of Primary Industries and the University of New England

and NRM (McMillan *et al.* 2017). This paper investigated the impact on covariance matrix estimates of including all SNPs in the same GRM or fitting separate GRMs for regular random SNPs and selected SNPs. The ratio of genetic variance explained by each genetic random effect was investigated, considering trait specific values of λ. Changes in covariances between genetic effects were also investigated.

MATERIALS AND METHODS

Data on reference flock animals from both the Sheep CRC Information Nucleus Flock (van der Werf *et al.* 2010) and MLA Resource Flock databases were obtained from the LAMBPLAN terminal sire analysis. Pre-adjusted phenotypic data were used for five traits: post-weaning weight (PWT, kg), carcase eye muscle depth (CEMD, mm), carcase c-site fat (CCFAT, mm), intramuscular fat (IMF, %) and shear force (SF5, Newtons). Phenotypes were only retained for animals with genotypes and where a phenotype was recorded for all six traits, resulting in 9688 animals with data. Phenotypes used were pre-adjusted for birth type, rearing type, age of measurement, age of dam, and hot carcase weight (trait dependant). Contemporary groups were taken from the LAMBPLAN analysis, with PWT belonging to one contemporary grouping (based on breed, flock, management group and sex, $n = 444$) and all carcase traits using different contemporary groupings (based on combinations of breed, flock, management group, sex and kill group, $n = 376$).

The 60410 SNPs available were split into two sets: the random SNPs ($n = 55{,}709$) and the predictive SNPs ($n = 4{,}701$). Three marker sets were then used to construct breed-adjusted genomic relationship matrices (GRMs), using the method described by Gurman *et al.* (2019). These GRMs were labelled $\boldsymbol{G}_r$ for the random SNPs, $\boldsymbol{G}_p$ for the predictive SNPs and $\boldsymbol{G}_{rp}$ for the combined set of SNPs. A corresponding pedigree-based relationship matrix for animals with genotypes was also constructed based on the extended pedigree including all known ancestors. To accommodate variance component estimation using the software package 'mtg2' (Lee *et al.* 2016), animal by animal relationship matrices were constructed for the other random effects to be considered, genetic groups and dam permanent environment. Genetic groups ($n = 89$) were included by constructing a matrix of pedigree-based breed proportions, $\boldsymbol{Q}$, where the rows sum to unity and animals with known parents are the average of their parental group proportions. These proportions were then converted to an animal by animal matrix by $\boldsymbol{QQ}^T$. Similarly, for the dam permanent environment effect, an incidence matrix was constructed relating dams to animals, $\boldsymbol{W}$, which was converted to an animal by animal matrix $\boldsymbol{WW}^T$.

Pairwise bivariate models for all trait combinations were then analysed using various combinations of the genetic random effect matrices described above. The general model fitted was $\boldsymbol{Y} = \boldsymbol{Xb} + \sum_{i=1}^{n} \boldsymbol{Zu}_i + \boldsymbol{e}$ where $\boldsymbol{Y}$ is the data in multivariate form; $\boldsymbol{X}$ is the incidence matrix for the contemporary groups; $\boldsymbol{b}$ is the vector of fixed-effect solutions; $\boldsymbol{Z}$ is the incidence matrix relating animals to breeding value estimates; $\boldsymbol{u}_i$ is the vector of random effect solutions for the ith random effect and $\boldsymbol{e}$ represents the residual. The model is also such that $\mathrm{var}(\boldsymbol{Zu}_i) = \boldsymbol{G}_i \otimes \boldsymbol{\Sigma}_i^2$ where $\boldsymbol{G}$ is the random effect matrix for the ith effect ($\boldsymbol{G} = \{\boldsymbol{A}, \boldsymbol{G}_r, \boldsymbol{G}_p, \boldsymbol{G}_{rp}, \boldsymbol{QQ}^T, \boldsymbol{WW}^T\}$ and $\boldsymbol{\Sigma}_i^2$ is the estimated covariance matrix for the random effect. For all models presented, genetic group and permanent environment effects of the dam were also included.

RESULTS AND DISCUSSION

Significantly higher log-likelihood values were found for the models that contained two GRMs. Models that included GRMs had higher heritabilities than the pedigree-only models (Table 1). Further, the highest trait heritabilities were observed in the models that contained $\boldsymbol{G}_{rp}$. The proportion of the total genetic variance explained by the GRMs was between 0.59 and 0.86, with the model containing $\boldsymbol{G}_{rp}$ explaining a slightly higher proportion than the model containing only $\boldsymbol{G}_r$

(Figure 1). The model that contained separate $\boldsymbol{G}_r$ and $\boldsymbol{G}_p$ either explained less variance than the model containing only $\boldsymbol{G}_r$ (see CCFAT and PWT) or less than the model containing $\boldsymbol{G}_{rp}$ (see CEMD, IMF and SF5). These estimates of λ are larger than the value of λ currently used in MERINOSELECT and LAMPLAN analyses, suggesting that further investigation is required to determine if this finding is consistent for other traits or if λ should be trait specific.

Table 1. Heritabilities calculated from the sum of all genetic effects in each model

Random Effect Model	PWT	CEMD	CCFAT	IMF	SF5
$\boldsymbol{A}$	0.217	0.202	0.225	0.629	0.305
$\boldsymbol{A} + \boldsymbol{G_r}$	0.283	0.225	0.252	0.636	0.313
$\boldsymbol{A} + \boldsymbol{G_{rp}}$	0.290	0.237	0.259	0.631	0.307
$\boldsymbol{A} + \boldsymbol{G_r} + \boldsymbol{G_p}$	0.274	0.227	0.253	0.614	0.267

Abbreviations: **A**: NRM, $\boldsymbol{G}_r$ GRM calculated from random SNPs, $\boldsymbol{G}_p$ GRM calculated from the predictive SNPs, $\boldsymbol{G}_{rp}$ GRM calculated from all SNPs

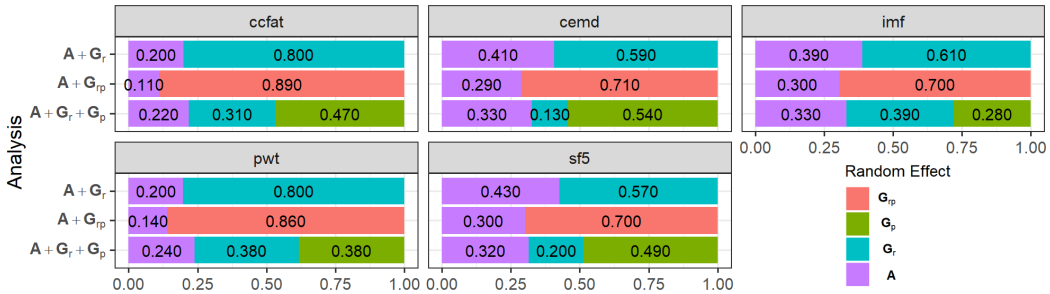

Figure 1. Proportions of the total genetic variance explained by each random effect. Abbreviations listed in Table 1

The genetic correlations between traits were not uniform across alternative models for genetic effects (Figure 2). While for most traits the correlations were fairly consistent, some trait pairs show much larger differences in the genetic correlations between models and random effects included, which the most evident of these being those correlations being CCFAT-PWT and CEMD-PWT. For both of these trait pairs, the estimated correlation was slightly negative between CCFAT-PWT and close to zero for CEMD-PWT from the model with only the NRM. When GRMs were added, these NRM correlations were estimated as strongly positive and the GRM correlations strongly negative. It should be noted that these differences largely cancel out when considering the overall genetic correlation. In some cases (CEMD-PWT, CF5-PWT, CEMD-SF5), the correlation estimated for the effects of $\boldsymbol{G}_r$ and $\boldsymbol{G}_p$ were different, suggesting here that the selected and random SNPs are capturing different genetic effects on these traits. Further investigation is required to determine why these differences in correlations occur.

A cross-validation study using the variance components from this study was also conducted to investigate the benefits on predictive ability of using two GRMs or a single GRM with all SNPs together in a large scale BLUP analysis (Li *et al.* 2021).

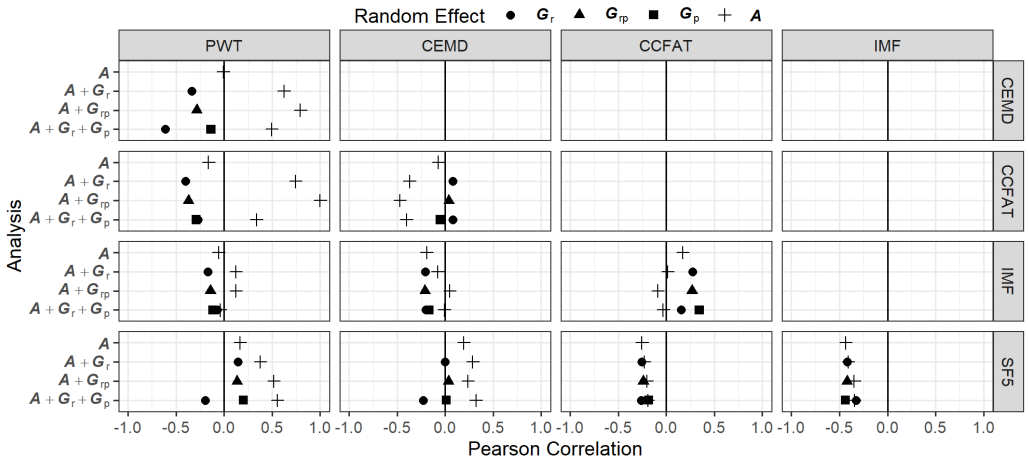

Figure 2. Estimated correlations between traits for the genetic random effects for all four models. Abbreviations listed in Table 1. One correlation was estimated to be larger than one and was therefore modified to one for presentation

CONCLUSIONS

This study found that the current value of $\lambda = 0.5$ used in Australian sheep genetic evaluations was lower than that estimated for the carcase and eating quality traits examined. Higher log-likelihoods values were estimated for the models containing two GRMs, however, this often resulted in slightly lower heritabilities compared to a model that contained all SNPs in one GRM. Including GRMs in the analysis resulted in different genetic correlations for some trait pairs from different GRM/NRM combinations. These results suggest that not considering the GRM in variance component estimation for SS-GBLUP can result in variances incorrectly proportioned between NRM and GRM. Further work is required to examine these impacts in other populations with different genomic population structures and in different traits.

ACKNOWLEDGEMENTS

This work was funded by MLA Project L.GEN. 1815.

REFERENCES

Brown D.J., Swan A.A., Boerner V., Li L., Gurman P.M., McMillan A., Van der Werf J.H.J. (2018) *Proc. World Congr. Genet. Appl. Livest. Prod*. **Species-Ovine**:460.

Gurman P.M., Bunter K.L., Boerner V., Swan A.A., Brown D.J. (2019) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **23**:254.

Lee S.H., Van der Werf J.H.J. (2016) *Bioinformatics* **32**:1420.

Li L., Gurman P.M., Swan A.A., Moghaddar N., Van der Werf J.H.J. (2021) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **24**:.*These proceedings*

McMillan A.J., Swan A.A. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **22**:557.

Moghaddar N., Khansefid M., Van der Werf J.H.J., Bolormaa S., Duijvesteijn N., Clark S.A., Swan A.A., Daetwyler H.D., MacLeod I.M. (2019) *Genet. Sel. Evol.* **51**:72.

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

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

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E., Visscher P.M. (2010) *Nat. Genet.* **42**:565.