Proc. Assoc. Advmt. Anim. Breed. Genet. 25: 213-216

ALIGNMENT OF PEDIGREE AND GENOMIC INFORMATION IN A COMBINED TERMINAL SIRE AND MATERNAL SHEEP ANALYSIS

P.M. Gurman, P. Alexandri, D.J. Brown, K.L. Bunter, S. de las Heras-Saldana, K.P. Gore, L. Li, A.J. McMillan, S.F. Walkom and A.A. Swan

Animal Genetics Breeding Unit[*], University of New England, Armidale, NSW, 2351 Australia

SUMMARY

Recent efforts to combine the Sheep Genetics LAMBPLAN maternal and terminal sire analyses presented the opportunity to review the genomic model used in single-step genomic BLUP. This paper uses forward validation to examine the suitability of different genomic models, including metafounders, a genomic relationship matrix (GRM) constructed from a single set of allele frequencies and a breed-adjusted GRM using allele frequencies for each identified subpopulation. The metafounder model yielded slightly higher stability values and was slightly less biased than the other genomic models. Metafounders also largely resulted in higher within-breed stabilities. These results support the use of a metafounder model for the combined LAMBPLAN analysis.

INTRODUCTION

The Australian sheep industry utilises diverse sheep breeds for their wool production, maternal abilities, along with meat and wool shedding properties. The genetic analysis for these sheep, LAMBPLAN, currently splits these animals into two groups, maternal dual purpose sheep breeds (primarily based on their mothering abilities) and terminal sire breeds (based mostly on their carcase qualities, Walkom *et al.* 2025), although the rise of composite sheep means breeds are often included in both analyses (McMillan *et al.* 2023). Subpopulations in these analyses have large genetic differences (Alexandri *et al.* 2025), with some subpopulations performing selection within their own population to maintain purity, while others are characterised by a diverse selection across multiple breeds, either to create a new stabilised subpopulation, or to target specific traits.

Single-step analyses incorporating multiple breeds or distinct subpopulations can be challenging due to the misalignment between the pedigree and genomic information. This is due to the numerator relationship matrix (NRM) coefficients ignoring breed differences, which are included in the genomic relationship matrix. This study examines approaches for the alignment of relationship matrices in a combined terminal-maternal sheep dataset with large breed structures present and aims to determine the model with the best predictive ability. The first of these, currently implemented in LAMBPLAN (Gurman *et al.* 2019), removes breed structure from the genomic relationship matrix, to allow alignment of the genomic relationship matrix and the pedigree relationship matrix (Makgahlela *et al.* 2013; Gurman *et al.* 2019). Metafounders (Legarra *et al.* 2015) provides a method to align the pedigree information to the genomics by adding population structure to the NRM.

MATERIALS AND METHODS

The data used here are described by Walkom *et al.* (2025) and the fixed effects component is described by de las Heras-Saldana *et al.* (2025). The pedigree included 4,839,523 animals, with 246,093 genotyped for 62k imputed SNPs, and 23 traits, with 3,658,685 animals being recorded for at least one trait. Multiple BLUP analyses compared different methods for aligning the genomic and pedigree relationship matrices. The first model was a pedigree-only analysis with genetic groups. The genetic groups covered 35 breed groups and up to 3-time windows. The second model was a

---

[*] A joint venture of NSW Department of Primary Industries and Regional Development and the University of New England

single-step model with a single set of scalar adjustments (Christensen 2012) based on the GRM constructed by Yang *et al.* (2010). The third method was to adjust the GRM for each animal's expected allele frequency given the estimated breed proportions from 27 subpopulation allele frequencies (Makgahlela *et al.* 2013; Gurman *et al.* 2019; Alexandri *et al.* 2025). This third method is currently used in the LAMBPLAN evaluations. The fourth and fifth models were based on transferring the genetic group definitions to metafounders. The matrix of relatedness between and within breeds, Γ, was estimated in two steps. Metafounders covering multiple time groups were merged and Γ was estimated using the back pedigree of the genotyped animals (Legarra *et al.* 2024a). This matrix was converted to F_{st} values (Legarra *et al.* 2024b). The collapsed Γ matrix was then expanded to cover time-based groups (Legarra *et al.* 2024a) based on pedigree inbreeding coefficients. When groups had insufficient genotypes (as observed by low Γ values during estimation of the merged Γ) the values in Γ for the base group were assumed for these metafounders. Metafounder analyses were performed as both pedigree only and genomic models assuming allele frequencies of 0.5 for all SNPs. All genomic analyses were performed using ssGTBLUP, (Mäntysaari *et al.* 2017) and the AGBU commercial solver.

Forward cross validation evaluated the most performant model. Animals born during or after 2021 were used as the validation (removing data for 716,842 animals, or 19.6% of animals with one or more phenotypes), with their phenotypes removed from the part analyses. Pairwise BLUP analyses were performed with (w) and without (p) the validation phenotypes, with the subsequent EBVs for the validation animals labelled as $\boldsymbol{u}_w$ and $\boldsymbol{u}_p$, respectively. Pairwise analyses were performed for each BLUP analysis described in the previous paragraph. The metrics of Legarra *et al.* (2018) were used to evaluate each model. We define here stability as $cor(\boldsymbol{u}_p, \boldsymbol{u}_w)$, which gives a relative indication of the model accuracy. Dispersion was defined as $\frac{cov(\boldsymbol{u}_p,\boldsymbol{u}_w)}{var(\boldsymbol{u}_p)}$, with values less than one indicating overdispersion and bias was defined as $\frac{\overline{u_p}-\overline{u_w}}{\sqrt{v_a}}$ where v_a is the genetic variance.

RESULTS AND DISCUSSION

A metric of genetic similarity, F_{st}, calculated from Γ with the collapsed time groups, is presented in Figure 1, grouped based on the breed types (maternal, terminal, wool, shedding and other). F_{st} values between some subpopulations were large, with the Border Leicester showing genetic differences from several other subpopulations. Differences were also observed between East Friesian and other groups but is likely due to low representation of East Friesians (see Figure 1 for Q matrix column sums for genotyped animals, considering crossbred contributions). Similar F_{st} ranges have been observed between some beef breeds (Manzari *et al.* 2024).

A summary of the cross-validation results across all subpopulations is presented in Table 1, and the stabilities for selected subpopulations are presented in Figure 2. The highest stabilities were observed for the genomic metafounder model. The genomic metafounder model also resulted in the least dispersion error and the least bias compared to the other genomic models. The stabilities within subpopulations show that for most subpopulations, the metafounder model was slightly more accurate than the other genomic models. Adding metafounders to the pedigree model also resulted in better predictive ability (Table 1), even with the pedigree model including genetic groups based on the same mapping of animals to groups/metafounders. One downside of the metafounders model was the higher iteration count and convergence time compared to the breed-adjusted GRM, but this reduced performance could be accepted given the gain in the predictive ability of EBVs.

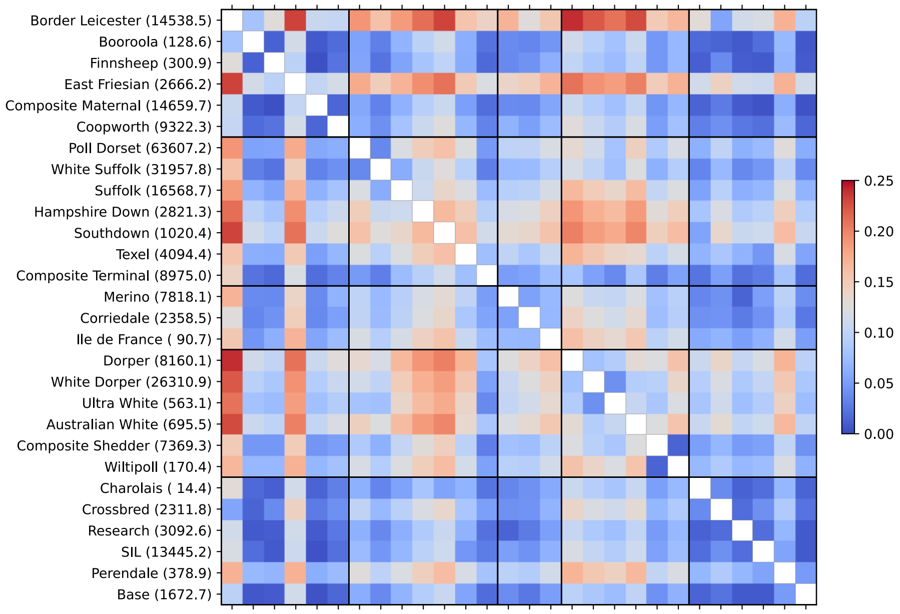

Figure 1. F_{st} value for each subpopulation calculated from the estimated gamma matrix. Black boxes group the subpopulations based maternal, terminal sire, wool, shedding and other subpopulations. Numbers next to each breed are the sums of the Q matrix columns for genotyped animals, considering contributions from crossbreds

Based on this research, metafounders have recently been incorporated into the routine BLUP evaluation pipeline for the combined LAMBPLAN analysis. This improvement results in changes to the routine pipeline including new parameters that need routine updates. For instance, the question of how often the gamma matrix should be re-estimated is unknown, but likely has similarities to how often the allele frequencies need to be re-estimated. Implementing metafounders will also require changes to the accuracy approximation algorithm. Bermann *et al.* (2023) has presented the differences in terms of true accuracies, with approximations requiring further research. Finally, the interpretation of relationships between animals diverges greatly from traditional interpretations (due to metafounder relationships) and changes to diagnostic information will need to be considered. While these are challenges to overcome, changing to a metafounder model will result in more accurate breeding values, and will allow for better across population selection decisions.

Table 1. Cross validation results and iteration counts and solving time for each model

	Stability	Dispersion	Bias	Iterations	Solve Time (hh:mm)
Pedigree	0.758	0.945	0.006	2,294	2:12
Pedigree with metafounders	0.776	0.961	0.020	3,143	3:29
Genomic with metafounders	0.851	0.944	0.025	3,684	7:05
Yang GRM	0.839	0.943	0.064	4,340	8:20
Breed Adjusted GRM	0.824	0.935	0.072	2,809	5:02

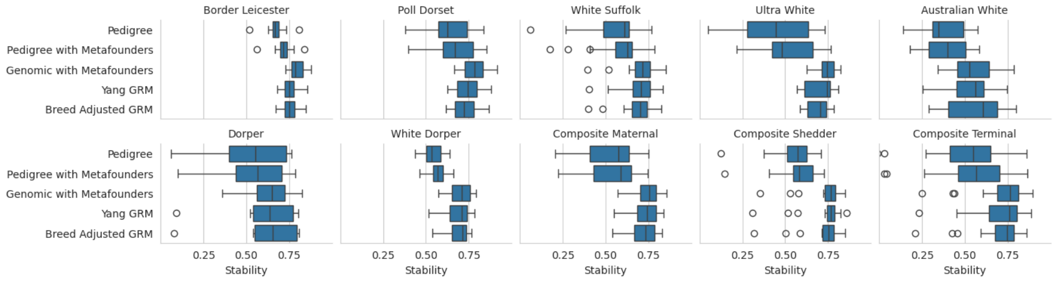

Figure 2. Cross validation stabilities as analysed within subpopulations

CONCLUSION

Metafounders improved predictive ability compared to the other genomic models considered. This work will power a future Combined LAMBPLAN analyses to provide breeders with more accurate breeding values and better across-breed predictions. Further work is required to fully document the benefits and to understand the ramifications of this change in a commercial context.

ACKNOWLEDGEMENTS

This research was funded by Meat and Livestock Australia project L.GEN.2204. The authors acknowledge the teams behind the Sheep CRC Information Nucleus and MLA Resource flocks as well as individual industry flock owners involved in flock management and data collection.

REFERENCES

Alexandri P., Brown D.J., Bunter K.L., De las Heras-Saldana S., Gore K.P., Gurman P.M., Li L., McMillan A.J., Walkom S.F. and Swan A.A. (2025) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **26**: *These proceedings*.

Bermann M., Aguilar I., Lourenco D., Misztal I. and Legarra A. (2023) *Genet. Sel. Evol.* **55**: 6.

Christensen O.F. (2012) *Genet. Sel. Evol.* **44**: 37.

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

De las Heras-Saldana S., Alexandri P., Brown D.J., Bunter K.L., Gore K.P., Gurman P.M., Li L., McMillan A.J., Walkom S.F. and Swan A.A. (2025) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **26**: *These proceedings*.

Legarra A., Bermann M., Mei Q. and Christensen O.F. (2024a) *Genet. Sel. Evol.* **56**: 1.

Legarra A., Bermann M., Mei Q. and Christensen O.F. (2024b) *Genet. Sel. Evol.* **56**: 34.

Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I. and Misztal I. (2015) *Genetics* **200**: 455.

Legarra A. and Reverter A. (2018) *Genet. Sel. Evol.* **50**: 53.

Makgahlela M.L., Strandén I., Nielsen U.S., Sillanpää M.J. and Mäntysaari E.A. (2013) *J. Dairy Sci.* **96**: 5364.

Mäntysaari E.A., Evans R.D. and Strandén I. (2017) *J. Anim. Sci.* **95**: 4728.

Manzari Z., Connors N.K., van der Werf J.H.J., Johnston D.J. and Ferdosi M.H. (2024). *Int Plant Anim. Geno. Conf.* **P007**

McMillan A.J., Walkom S.F. and Brown D.J. (2023) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **25**: 31.

Walkom S.F., Alexandri P., Brown D.J., Bunter K.L., De las Heras-Saldana S., Gore K.P., Gurman P.M., Li L., McMillan A.J. and Swan A.A. (2025) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **26**: *These proceedings*.