Proc. Assoc. Advmt. Anim. Breed. Genet. 25: 158 - 161

ASSESSING THE VALUE OF METAFOUNDERS FOR GENOMIC PREDICTION IN AUSTRALIAN SIMMENTAL BEEF CATTLE

D.J.A. Santos, N.K. Connors, P.M. Gurman, M.H. Ferdosi, S.P. Miller and A.A. Swan

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

SUMMARY

The “metafounders” framework is used to augment relationship matrixes to accommodate genetic structure in founder populations, and can be estimated from genotypes, making it useful to align pedigree and genomic relationships in single-step genomic analyses. This paper aimed to assess the value of metafounders in the genomic evaluation of beef traits in Australian Simmental cattle, and in particular the possibility of collapsing genetic groupings based on metafounder similarity. Estimated breeding values from metafounder models with different groupings had similar predictive ability across 12 beef traits, while models with higher weighting on genomic relative to pedigree information tended to perform better.

INTRODUCTION

Metafounders (MF) are pseudo-individuals included in the pedigree that allow accounting for genetic heterozygosity and relationships within and between base populations, considering unknown ancestral populations (Legarra *et al.* 2015). The MF approach may be advantageous because it derives compatibility between genomic (**G**) and pedigree (**A**) relationship matrices by modifying **A** to align with **G** (Garcia-Baccino *et al.* 2017). Currently, the BREEDPLAN genetic evaluation for Australian Simmental uses 25 genetic groups, defined based on the country of origin, breed, and year of birth of animals with unknown parentage. The influence of all these genetic groupings and structures in the pedigree of Australian Simmental need to be considered in single-step genetic evaluations. This study aimed to assess the utility of MF in the genomic evaluation of beef traits in Australian Simmental, considering the predictive ability with different MF assignment strategies in the pedigree.

MATERIALS AND METHODS

Data. The genomic data consisted of 8,245 genotyped animals with 59,678 SNPs. Traits analysed included eight live ultrasound scan body composition traits, eye muscle area, intramuscular fat, P8 fat, and rib fat in bulls and heifers (BEA, BIM, BP8, BRF, HEA, HIM, HP8 and HRF), and four body weight traits, birth (BWD), weaning (WWD), yearling (YWD), and final weight (FWD). Numbers of genotyped and pedigree-only animals recorded for each trait are shown in Table 1.

MF procedures. Metafounders were included in single-step models using an adapted inverse relationship matrix defined as $\mathbf{H}^{\Gamma-1} = \mathbf{A}^{\Gamma-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{\Gamma-1} \end{bmatrix}$, where $\mathbf{A}^{\Gamma}$ is the pedigree relationship matrix augmented by the “gamma” matrix modelling within and across base population relationships, $\mathbf{A}_{22}^{\Gamma}$ is the sub-matrix of $\mathbf{A}^{\Gamma}$ for genotyped animals, and $\boldsymbol{\Gamma}$ is the gamma matrix (Legarra *et al.* 2015). The matrix **G** was obtained as $\lambda\mathbf{G}_m + (1 - \boldsymbol{\lambda})\mathbf{A}_{22}^{\Gamma}$, where $\mathbf{G}_m$is the genomic relationship matrix as calculated via VanRaden (2008), and λ is the weighting factor between genomic and pedigree relationship matrices, set as either 0.5 or 1. For λ=1 a small positive value was added to the diagonal of $\mathbf{G}_m$ to ensure it was invertible.

As described above, the genetic groups used to define MF groups have been defined based on country of origin, breed, and year of birth of animals with unknown parents. In addition to 12

---

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

Simmental groups, there were 4 substantial groups of Angus origin, with minor contributions from Hereford, European, Indicus, and unrecognised breeds.

Estimation of **Γ** was performed via generalised least squares (Garcia-Baccino *et al.* 2017) using the WOMBAT software package (Meyer 2007). The correlations between MF were calculated from the **Γ** estimated, and the MF were grouped using multivariate clustering techniques.

Genetic evaluation and prediction design. Prior to cross-validation, the variance components for each trait were estimated using all data available with the WOMBAT program (Meyer 2007), and these parameters used for BLUP analyses. Variance component estimation and EBV predictions using ssGBLUP with MF were performed using a single-trait animal model with contemporary groups as a fixed effect, direct genetic effects fitted as random for all traits, and maternal genetic effects fitted for BWD and WWD only (uncorrelated with direct genetic effects). Phenotypes were pre-adjusted for fixed effects apart from contemporary group.

The performance of analyses with different MF groupings was compared across traits using a k-fold cross-validation approach with k=5. For the k-folds analyses, animals with phenotypic and genotypic data were randomly split into five parts. EBVs were calculated 5 times for each trait, omitting the phenotypes of animals in each validation set such that their EBVs were then predicted from genomic and pedigree relationships ("part" EBVs). Then, the accuracy, stability and dispersion of the predictions were assessed. Accuracy was calculated as the correlation between part EBV and phenotypes of validation animals for all traits except for the two maternally influenced traits (BWD and WWD) for which the LR method was used (Legarra and Reverter, 2018). Stability was calculated as the correlation between part and full EBVs for the validation animals, and bias as the regression of part EBVs on phenotypes. Results for each statistic were averaged across folds.

RESULTS AND DISCUSSION

Gamma matrix and MF clustering. The matrix **Γ** was estimated for 25 MFs (MF25) and the correlations between MFs grouped by similarity are shown in Figure 1. The diagonal "self-relationship" elements of **Γ** ranged from 0.29 to 0.82 with an average of 0.47 (the possible range in values is 0 to 2 with higher values indicating higher inbreeding). The average for Simmental groups was 0.43 and for Angus was 0.56. Higher values tended to be for smaller groups which by default have less diversity. Corresponding ancestral correlations were typically >0.8 within the Simmental and Angus groups, and approximately 0.2 to 0.6 between other groups (Figure 1).

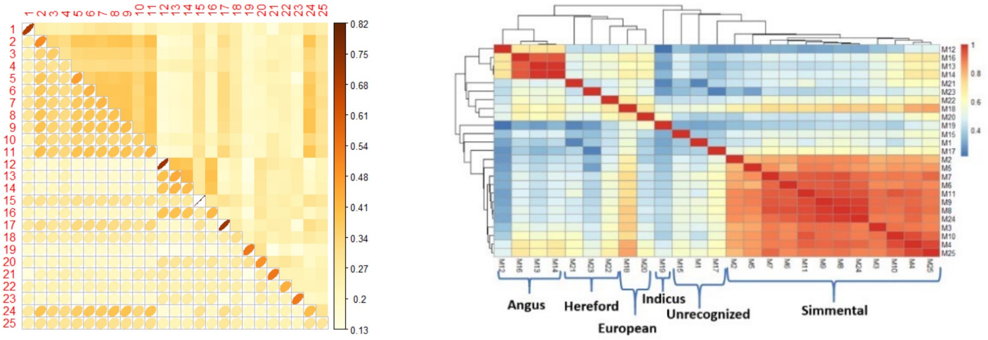

Figure 1. Metafounder clustering results: top left = Gamma matrix (*Γ*) estimated for 25 genetic groups, top right = Gamma correlation matrix with clustering and dendrogram of the genetic groups

Through a k-means algorithm, the MF were collapsed progressively into 15, 14 and 12 clusters, and new $\mathbf{\Gamma}$ matrices estimated. In all cases the 3 most similar groups of Angus origin were collapsed, while 12 Simmental groups were collapsed into 4 groups in MF15, 3 groups in MF14, and 1 group in MF12.

Genetic Parameters. Heritability estimates from the MF25 models are shown in Table 1. Estimates for MF12, 14, and 15 were very similar to MF25 and are therefore not shown. These results are similar to the heritabilities assumed in the BREEDPLAN analysis for the breed, although generally marginally higher. According to Legarra *et al.* (2015), genetic variance estimates obtained from MF models should not be interpreted as a genetic variance within the population but as a parameter of the statistical model used for the analysis. Heritability estimates tended to be higher for models with $\lambda = 0.5$.

Table 1.Number of genotyped (Geno) and pedigree only animals (Ped) with records for each trait, and heritability estimates for MF25 models with λ=1 or 0.5

Trait	Geno	Ped	Heritability (λ=1)	Heritability (λ=0.5)
BEA	1,800	21,017	0.32	0.33
BIM	1,680	11,339	0.28	0.28
BP8	1,796	20,986	0.37	0.42
BRF	1,795	20,889	0.28	0.31
HEA	483	15,787	0.35	0.36
HIM	482	9,417	0.42	0.42
HP8	479	15,759	0.56	0.57
HRF	476	15,746	0.47	0.48
BWD	3,068	111,262	0.40	0.40
WWD	2,786	115,209	0.27	0.40
YWD	2,842	118,646	0.42	0.42
FWD	1,647	64,860	0.44	0.45

Cross-validation. Accuracies across traits for MF models with $\lambda = 0.5$ and 1 are shown in Figure 2. There was no effect on accuracy for analyses with different MF groupings, but an increase in accuracy was observed with $\lambda = 1$ for body weight traits. This trend was not observed for body composition traits. Stability of part versus full EBVs (Figure 3) was also higher for models with $\lambda = 1$, but again there was no difference between MF groupings. Results for dispersion (not shown) were similar across models, and sufficiently close to the expected value of 1 across traits. These results suggest reasonable prediction accuracy can be obtained using MF models, with some evidence of higher accuracy with higher λ values. However, there was no advantage in aggregating groups based on similarity.

Before implementation, additional studies should be performed to compare these MF analyses with traditional genetic groups models, and to investigate the accuracy of estimating MF relationships for groups with low numbers of genotypes.

CONCLUSIONS

Although patterns of similarity between metafounder groups were evident, generally reflecting breed of origin, there was little apparent benefit in collapsing groups. Alternatively, simplification of groups may be possible if desired, providing the performance differences between groups to be collapsed are minimal.

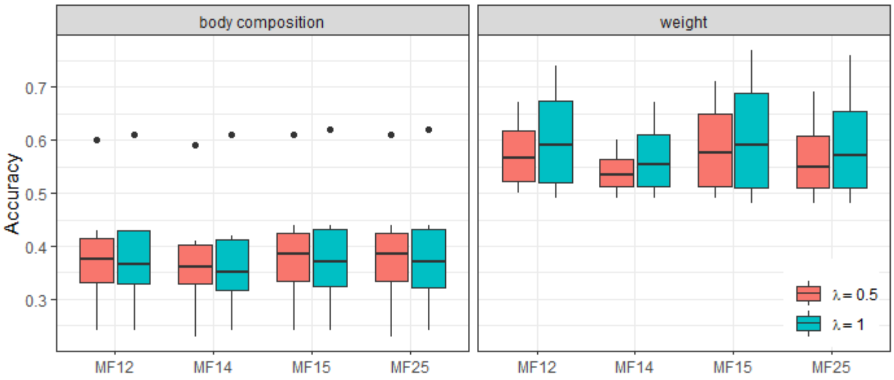

Figure 2. distribution of cross-validation accuracy across traits for MF models with λ=1 or 0.5

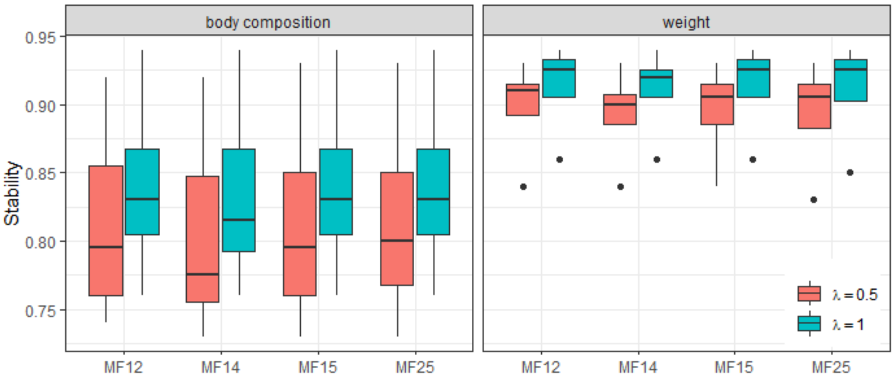

Figure 3. distribution of cross-validation stability of EBVs across traits for MF models with λ=1 or 0.5

ACKNOWLEDGEMENTS

This project was supported by Meat and Livestock Australia L.GEN.2204. The authors thank Simmental Australia for providing access to their data and the Agricultural Business Research Institute for providing the data extracts.

REFERENCES

Garcia-Baccino C. A., Legarra A., Christensen O. F., Misztal I., Pocrnic I., Vitezica Z. G. and Cantet R. J. C. (2017) *Genet. Sel. Evol.* **49:** 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.
Meyer K. (2007) *J. of Zhejiang Uni*. SCIENCE B **8:** 815.
VanRaden P.M. (2008) *J. Dairy Sci.* **91:** 4414.