

## **‘METAFOUNDERS’ TO MODEL BASE POPULATIONS IN GENOMIC EVALUATION FOR MULTI-BREED SHEEP DATA**

**K. Meyer and A.A. Swan**

Animal Genetics & Breeding Unit\*, University of New England, Armidale, NSW, 2351 Australia

### **SUMMARY**

Models for genetic evaluation of animals from different base populations need to account for systematic differences not explained by genetic relationships considered. These include differences between breeds, animals with unknown parentage born in different time periods or, for single-step evaluation, founders for animals with or without genotype information. A standard method to achieve this, is to define appropriate genetic groups and fit these as additional effects in the model of analysis. Recently, so-called meta-founders have been proposed as an alternative which accounts for ancestral inbreeding and relationships, estimated from genomic information. We examine estimates of ancestral relationships and their impact on predicted breeding values for a practical data set from a multi-breed sheep population. While estimates were afflicted by insufficient genomic information for some groups, results correctly identified some known breed or strain differences and patterns of introgression. Correlations between predicted breeding values from respective analyses fitting genetic groups and meta-founders were high, suggesting that there is scope for meta-founders to replace genetic groups. However, fitting meta-founders reduced variances of predicted breeding values. Further investigations when more genotype information becomes available are warranted.

### **INTRODUCTION**

The single-step procedure for joint genetic evaluation of genotyped and non-genotyped animals using both pedigree and genomic information has become routine for many livestock improvement schemes. Commonly, this is implemented as the so-called ssGBLUP which replaces the classic, pedigree based relationship matrix, **A**, with its counterpart, **H**, which combines the genomic relationship matrix, **G**, with **A**. An inherent problem with this approach is that **A** and **G** imply conceptually different base population: For **A**, parents of animals at the time when pedigree recording began are considered to be the unrelated, non-inbred founders. In contrast, genomic relationships reference an ancestral base population in the distant past. Several methods have been described to align the two matrices; see Meyer *et al.* (2018) for a recent review. Some proposals involve scaling **G** to ‘match’ **A** while others suggest to modify **A** to account for ancestral inbreeding (Christensen 2012). Specifically, the latter can be achieved by replacing unknown parents in the pedigree with ‘meta-founders’ (MF), allowing for ancestral inbreeding and relationships between them, estimated from genomic information (Legarra *et al.* 2015; Garcia-Baccino *et al.* 2017). MF are conceptually similar to the ‘phantom’ parents (Westell *et al.* 1988) used routinely to account for unknown parent groups. Thus, in addition to aligning **G** and **A**, they may provide an alternative to modeling genetic groups. Moreover, MF are treated as correlated and may model genetic relationships between different base populations more appropriately. This paper examines estimates of ancestral relationships and their effects on estimates of breeding values for a practical data set recorded for Australian sheep.

### **MATERIAL AND METHODS**

Data consisted of 1,206,908 records for eye muscle depth, recorded for Australian terminal sire sheep breeds between 1990 and 2018. These included 5 main breeds, namely Poll Dorset, Suffolk, White Suffolk, Merino and Texel, and 17 minor breeds with breed differences modeled by appropriately defined genetic group effects. After eliminating individuals not connected to the data or

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

genotyped animals, there were 1,698,838 animals in the pedigree. Genotype information, consisting of marker counts for 48,599 SNPs, was available for 23,040 animals, of which 18,396 had phenotypes. Data were pre-corrected for fixed effects of birth and rearing type, age, dam age and body weight.

Routine analyses currently classify animals of unknown parentage into 93 genetic groups (GG), based on flock and year of birth. These animals were assigned MF ‘parents’ based on GG memberships. A total of 10.6% of animals had both parents unknown and 7.8% had no sire identified. All animals with both parents unknown belonged to a single GG thus had the same MF as ‘sire’ and ‘dam’.

Estimates of the matrix of ancestral relationships,  $\mathbf{\Gamma}$ , were obtained from marker information using a pseudo-EM algorithm (Garcia-Baccino *et al.* 2017; Legarra and Astruc 2018). For two of the GG, no genotypes were available. For these, diagonal elements of  $\mathbf{\Gamma}$  were set to the minimum value found for the other groups. Similarly, off-diagonal elements were replaced by values reflecting the minimum correlation encountered. In addition, the resulting estimate of  $\mathbf{\Gamma}$  was regularised by shrinking its eigenvalues towards their mean, so that the smallest value exceeded 0.01. The inverse numerator relationship matrix including MF,  $\mathbf{A}^{-\Gamma}$ , and the corresponding submatrix of  $\mathbf{A}$  for genotyped animals,  $\mathbf{A}_{22}^{\Gamma}$ , were obtained as outlined by Legarra *et al.* (2015).

A ‘raw’ genomic relationship matrix,  $\mathbf{G}_M$ , was build from marker counts using method 1 of Van Raden (2008). This was transformed into  $\mathbf{G} = \lambda(\mathbf{G}_M + \alpha\mathbf{J}) + (1 - \lambda)\mathbf{A}_{22}$  with  $\lambda = 0.95$  and  $\mathbf{A}_{22}$  the submatrix of  $\mathbf{A}$  for genotyped animals. To build the ‘standard’  $\mathbf{H}^{-1}$  (no MF) markers were centered using observed frequencies and  $\alpha = 0.025$  was estimated following Vitezica *et al.* (2011). To build  $\mathbf{H}^{-\Gamma}$  (including MF), markers were centered assuming allele frequencies of 0.5 and  $\alpha = 0$ , and  $\mathbf{A}^{-1}$  and  $\mathbf{A}_{22}$  were replaced by  $\mathbf{A}^{-\Gamma}$  and  $\mathbf{A}_{22}^{\Gamma}$ , respectively. In addition,  $\mathbf{H}^{-\Gamma}$  was scaled (see Legarra *et al.* 2015) so that the same variance components were appropriate for analyses with and without MF.

The model for ssGBLUP analyses fitted animals’ additive genetic effects, 54,094 contemporary groups (fixed) and 56,212 sire  $\times$  flock-year (random) effects throughout. A standard analysis (no MF) fitted 93 GG as additional random effects. For analyses with MF,  $\mathbf{H}^{-1}$  was replaced with  $\mathbf{H}^{-\Gamma}$  either including or excluding GG. Mixed model equations were solved iteratively using a preconditioned conjugate gradient algorithm with diagonal preconditioner. All calculations were performed using WOMBAT (Meyer 2007).

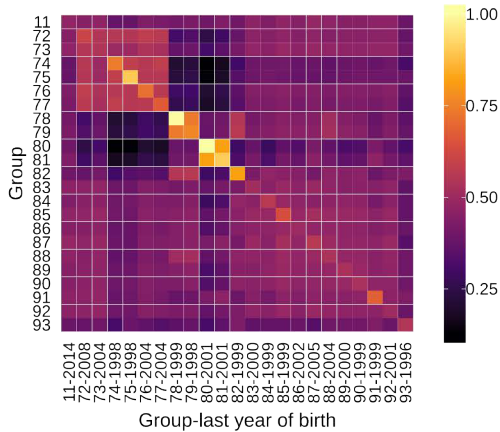
## RESULTS AND DISCUSSION

Means and ranges for estimates of ancestral inbreeding or ‘self-relationships’, i.e. the diagonal elements of  $\mathbf{\Gamma}$ , and correlations between MF (derived from  $\mathbf{\Gamma}$ ) are summarised in Table 1. Mean across breed group correlations ranged from 0.48 (Suffolk  $\times$  Texel) to 0.71 (Poll Dorset  $\times$  White Suffolk), with the range of individual values similar to that within the minor breeds and Merinos (0.14 to 0.95). Some correlations close to unity suggest scope for merging selected GG. Overall, however, estimates fluctuated considerably and no consistent breed group differences or time trends were evident. To some extent, this can be attributed to definitions of GG, e.g. multiple groups for the same breeds in different flocks and overlapping years of birth. In contrast, Legarra and Astruc (2018) found increasing inbreeding and covariances between MF with time for a breed of French dairy sheep. Accurate estimation of  $\mathbf{\Gamma}$  requires sufficient genomic information for all MF. Hence, in part at least, this variability

**Table 1. Estimates of self-relationships and ancestral correlations between meta-founders**

	n <sup>a</sup>	Self-relationship		Correlation	
		$\bar{x}^b$	range	$\bar{x}$	range
P. Dorset	15	0.68	0.57–0.89	0.83	0.53–0.98
Suffolk	14	0.88	0.57–1.07	0.61	0.37–0.87
W.Suffolk	14	0.62	0.50–0.81	0.75	0.58–0.98
Merino	23	0.66	0.48–1.00	0.67	0.14–0.95
Texel	10	0.80	0.59–0.96	0.62	0.40–0.96
Other	17	0.70	0.45–0.99	0.51	0.14–0.95

<sup>a</sup> No. of MFs per breed group    <sup>b</sup> Mean



**Figure 1. Heatmap plot of diagonal block of  $\Gamma$  for Merino groups**

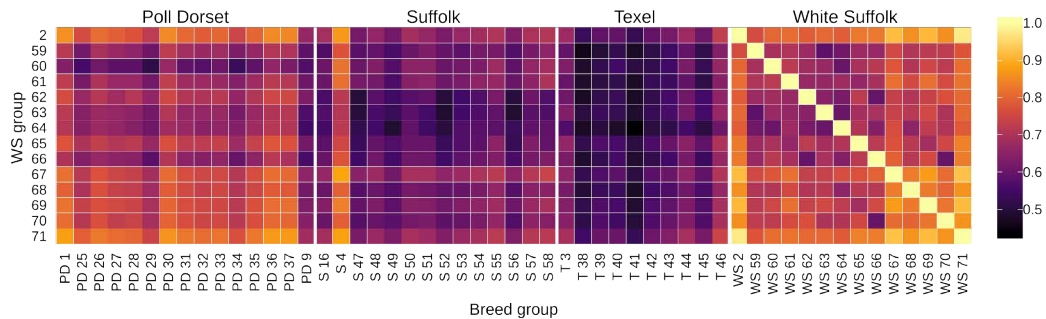
may be attributed to a rather uneven distribution of genotypes over GG and MF. Genotyped animals mostly contributed to numerous GG. Sums of relative contributions (0 to 1; summed over genotyped animals) were less than 5 for 23 GG and exceeded 500 for 7 GG.

Nevertheless, estimates of ancestral relationships correctly identified some known strain differences or patterns of introgression. Figure 1 shows the diagonal block of  $\Gamma$  for Merino GG. Groups 72–81 originate from a resource flock (e.g. Taylor and Atkins 1997), with 72–77, 78–79 and 80–81 representing medium, strong and fine wool selection lines, respectively, while 82 is another strong wool flock. The pattern of covariances between these GG reflects the divergence between strains. Similarly, high an-

cestral correlations between breed groups shown in Figure 2 highlight the role of Poll Dorset sheep in the formation of the White Suffolk breed.

Statistics comparing predicted breeding values (EBV) from analyses fitting GG or MF are summarised in Table 2 for different categories of animals. Overall, correlations were high suggesting that there is scope for MF to replace explicit GG in the model of analysis. Variances of EBV fitting MF only were considerably lower than those obtained fitting GG as an additional random effect. This implies somewhat stronger shrinkage of predictions when fitting MF, in spite of assuming the same variances for GG and animals' additive genetic effects, or, on average, higher error variances for MF than GG effects. The correlation between predicted GG and MF was 0.87. Negative intercepts highlight the change in alignment of conceptual base populations due to MF. As to be expected, fitting both MF and GG tended to increase these variances but had little effect on correlations. It also increased some of the regressions coefficients for EBV fitting MF on EBV fitting GG, presumably by accounting for group differences which were not quite modeled adequately, possibly due to lack of genomic information and thus less reliable estimates of ancestral relationships.

For routine implementation of ssGBLUP, convergence behaviour of iterative schemes to solve the mixed model equations is important. Fitting GG is known to increase the number of iterates required considerably. Replacing GG by MF did not prove advantageous in this respect: For our analysis fitting MF increased the number of iterates substantially, from 619 (GG) to 1,014 (MF). This



**Figure 2. Heatmap plot of estimates of ancestral correlations for White Suffolk groups**

**Table 2. Summary statistics for predicted breeding values from different analyses**

Gen. <sup>a</sup>	Ph. <sup>b</sup>	MF <sup>c</sup>	N <sup>d</sup>	Fit MF only				Fit MF and GG			
				V <sup>e</sup>	r <sup>f</sup>	b <sup>g</sup>	a <sup>h</sup>	V	r	b	a
No	No	0	204,688	0.724	0.991	0.887	-0.317	0.897	0.993	0.940	-0.229
No	Yes	0	1,159,639	0.816	0.996	0.912	-0.337	0.850	0.997	0.919	-0.220
No	No	1	103,126	0.517	0.978	0.836	-0.292	0.930	0.976	0.941	-0.227
No	Yes	1	28,873	0.554	0.992	0.824	-0.294	0.740	0.992	0.853	-0.174
No	No	2	179,472	0.421	0.976	0.840	-0.291	0.940	0.980	0.951	-0.218
Yes	No	0	4,562	0.916	0.989	0.937	-0.423	1.009	0.994	0.999	-0.459
Yes	Yes	0	18,341	0.900	0.995	0.937	-0.398	0.989	0.996	0.990	-0.436
Yes	No	1	82	0.815	0.981	0.898	-0.945	0.768	0.973	0.853	-0.570
Yes	Yes	1	55	0.838	0.970	0.884	-0.737	0.837	0.972	0.889	-0.582

<sup>a</sup> Genotype    <sup>b</sup> Phenotype    <sup>c</sup> Number of MF parents    <sup>d</sup> Number of animals    <sup>e</sup> Variance of predicted breeding values as proportion of variance fitting GG only    <sup>f</sup> Correlation with breeding values fitting GG only    <sup>g</sup> Regression on breeding values fitting GG only    <sup>h</sup> Intercept

can be attributed to allowing for correlations between groups when fitting MF. In particular, some correlation estimates were close to unity (see Table 1). Hence, convergence is likely to be improved if groups can be redefined so as to avoid small eigenvalues in  $\Gamma$ . Furthermore, additional analyses using a 'deflated' preconditioner (see Meyer and Swan 2019) reduced iterates needed when fitting MF to 594, suggesting that there is scope to compensate for any increases in numbers of iterates required due to fitting MF rather than GG.

## CONCLUSIONS

Meta-founders have been proposed to align base populations for pedigree based and genomic relationship matrices in ssGBLUP and as an alternative to modeling breeds or genetic groups. Results demonstrate that estimating ancestral relationships and fitting MF is feasible for practical data with many genetic groups. However, optimal performance requires careful definition of groups and sufficient genomic data for all groups to ensure reliable estimates of ancestral relationships.

## ACKNOWLEDGEMENTS

This work was supported by Meat and Livestock Australia grant L.GEN.1704.

## REFERENCES

- Christensen O.F. (2012) *Genet. Sel. Evol.* **44**:37.
- 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. and Astruc J.M. (2018) In *Book of Abstracts of the 69th Annual Meeting of the European Federation of Animal Science*. Dubrovnik, Croatia, 27-31 August, pp. 616, Paper no. 66.47.
- Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I. and Misztal I. (2015) *Genetics* **200**:455.
- Meyer K. (2007) *J. Zhejiang Univ. SCIENCE B* **8**:815.
- Meyer K. and Swan A. (2019) *Proc. Ass. Advan. Anim. Breed. Genet.* **23**:000.
- Meyer K., Tier B. and Swan A. (2018) *Genet. Sel. Evol.* **50**:39.
- Taylor P.J. and Atkins K.D. (1997) *Int. J. Sheep Wool Sci.* **45**:92.
- Van Raden P.M. (2008) *J. Dairy Sci.* **91**:4414.
- Vitezica Z.G., Aguilar I., Misztal I. and Legarra A. (2011) *Genet. Res.* **93**:357.
- Westell R.A., Quaas R.L. and Van Vleck L.D. (1988) *J. Dairy Sci.* **71**:1310.