ADJUSTING THE GENOMIC RELATIONSHIP MATRIX FOR BREED DIFFERENCES IN SINGLE STEP GENOMIC BLUP ANALYSES

P.M. Gurman, K.L. Bunter, V. Boerner, A.A. Swan and D.J. Brown

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

SUMMARY

The genomic relationship matrix (GRM) routinely constructed for single-step genomic BLUP analyses is known to contain breed structure, observable via principal component analysis, while the pedigree relationship matrix uses coefficients that are constant between known relatives regardless of breed or genetic group membership. This paper explores the effect of using allele frequencies for each breed or genetic group when calculating the GRM to reduce breed or genetic group structures in the GRM in the presence of pedigree based genetic groups fitted as random effects. We investigated the effect of using a breed-adjusted GRM on estimated breeding values, showing cross-validation results, genetic trends and estimated breeding value accuracies. Cross-validation results across breed showed a slight increase in EBV accuracy using a breed-adjusted GRM, 0.220 ± 0.068 compared to a non-adjusted GRM, 0.206 ± 0.071. Genetic trends calculated from estimated breeding values (EBVs) using a breed-adjusted GRM were more closely aligned to those estimated using a pedigree-only model compared to a non-adjusted GRM. These results show that using a single set of allele frequencies in a GRM with a diverse number of breeds can result in biased breeding values and biased genetic trends relative to those obtained from pedigree model including breed groups.

INTRODUCTION

With the transition of routine genetic evaluations from pedigree- or genomic blending-based approaches, to single-step (Legarra *et al.* 2014), the alignment of the GRM to the pedigree-based numerator relationship matrix (NRM) has become a focus of research interest when genetic groups are present and included in the model as separate random effects. This research focus is, in part, due to the impact that any misalignment can have on genetic trends (Meyer *et al.* 2018). Scalar adjustment parameters have been suggested (Vitezica *et al.* 2011; Christensen 2012) to align the NRM and GRM, while leaving the general structure inherent in the GRM intact. The 'metafounders' framework (Legarra *et al.* 2015; Garcia-Baccino *et al.* 2017) was suggested as a method for modifying the NRM to be in better alignment with the GRM and in doing so, replacing genetic groups (Westell *et al.* 1988) that are currently used for managing missing pedigree. While the metafounders framework is a promising method for handling misalignment of the NRM with the GRM and genetic groups, it is challenging to implement in routine analyses. That approach assumes that each metafounder has genotyped animals in their forward pedigree, which may not occur in practice, and may require modifications to the rules currently used to assign animals to metafounders or genetic groups. An alternative method is to align the GRM to the NRM by removing breed/genetic group structure from the GRM, as described by Makgahlela *et al.* (2013). This paper aims to examine the impact of using a breed-adjusted GRM (hereby BGRM) in routine single-step genomic BLUP analyses on cross-validation correlations and regression slopes of evaluations on adjusted phenotypes and genetic trends compared to a standard non-adjusted GRM (hereby SGRM).

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

MATERIALS AND METHODS

A SGRM using the method of Yang *et al.* (2010) can be constructed as $\boldsymbol{W} = \frac{\boldsymbol{M} - 2\boldsymbol{p}}{\sqrt{2\boldsymbol{p}(1-\boldsymbol{p})}}; \boldsymbol{G} = \frac{\boldsymbol{W}\boldsymbol{W}'}{m}$, where $\boldsymbol{M}$ is the marker matrix of dimensions animals by markers, $\boldsymbol{p}$ is the allele frequency of the animals in $\boldsymbol{M}$ and m is the number of markers in $\boldsymbol{M}$. The above equation was extended by Makgahlela *et al.* (2013) for allele frequencies that vary by breed (BGRM). A breed proportion matrix, $\boldsymbol{Q}$, was calculated from genotypes using BreedComp (Boerner *et al.* 2018), and the allele frequency for each breed in Q was calculated as F, allowing the method of Yang *et al.* (2010) to be extended for each column in the Q matrix. The expected allele frequency for each animal based on its breed proportion is then estimated as $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{F}$, and thus $\boldsymbol{W}_{ij} = \frac{\boldsymbol{M}_{ij} - 2\boldsymbol{P}_{ij}}{\sqrt{2\boldsymbol{P}_{ij}(1-\boldsymbol{P}_{ij})}}$ and $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{W}'/m$. To examine the differences in EBVs using a pedigree-only relationship matrix or these two GRM construction methods in single step analyses utilising pedigree based genetic groups implemented as random effects, multi-trait BLUP analyses were performed on maternal reproduction data from sheep. The EBVs (including genetic group estimates) from these analyses were compared via genetic trends and EBV correlations for animals born after 2013. The data consisted of approximately 2.4 million animals in the pedigree, with 11,761 of these genotyped and phenotypes collected on up to 15 traits. Reproduction traits that were included in the analysis were: fertility of yearling (ycon) and adult (con) ewes, litter size of yearling (yls) and adult (ls) ewes, rearing ability of yearling (yera) and adult (era) ewes, and maternal behaviour score of adult ewes (mbs). Other traits included in this analysis were: post-weaning eye muscle depth (pemd), post-weaning carcase fat (pcf), post-weaning scrotal circumference (psc), yearling scrotal circumference (ysc), pre-joining weight of post-weaning (pwt) and adult (awt) ewes, and pre-joining condition score of yearling (ycs) and adult (cs) ewes. The number of phenotypes per trait varied, ranging from 595,978 (ls) to 1,746 records (ycs). The genotypes represented a variety of breeds, dominated by Border Leicester, Coopworth, Corriedale and crossbred animals, including Border Leicester-Merino cross sheep. BLUP analyses were performed using each of these three relationship matrices assuming common variance components, and included random effects for genetic groups. Further model details can be found in Bunter et al. (2019). Forward cross-validation was performed. Phenotypes for animals born after 2013 were removed from the analysis and breeding values were estimated for these animals from the remaining phenotypes. This year of birth was chosen to ensure sufficient reproduction records were included in the validation set, though some traits still had few validation phenotypes. Phenotypes recorded after 2013 were then adjusted for the relevant fixed effects to calculate correlations with the estimated breeding values, with phenotypes re-scaled by the square root of the heritability. For each trait, adjusted phenotypes were regressed on the EBVs; slopes less than one indicate over-prediction (i.e. bias) and slopes above one indicate under-prediction. The mean and standard deviation of the correlations and regression slopes across all 15 traits was calculated, weighted by the number of animals with phenotypes included in the validation set. Traits with fewer than 300 observations (n=4) were not included in these means.

RESULTS AND DISCUSSION

Comparing the breeding values of all 15 traits for animals born after 2013 estimated using the SGRM with those from a pedigree only analysis, the mean correlation was 0.988. The minimum correlation was 0.977, while the maximum was 0.997. The same correlations calculated using a MGRM were 0.996, with a minimum correlation of 0.993 and a maximum of 0.999. Within Border Leicester sheep, the mean correlation for a SGRM and a MGRM changed from 0.973 to 0.992, respectively, within Coopworth sheep the mean correlation changed from 0.978 to 0.999 and within Corriedale sheep the correlation increased from 0.983 to 0.999. The genetic trends for the four traits

showing the lowest correlations between EBVs from pedigree and SGRM models are presented in Figure 1. These correlations and genetic trends show that using the BGRM produced breeding values and trends that were closer to those previously observed using the NRM and breed group effects.

The mean correlations from forward cross-validation estimated using NRM, SGRM and BGRM were 0.196 ± 0.0824 , 0.206 ± 0.071, and 0.220 ± 0.068, respectively. The mean regression slopes estimated using NRM, SGRM and BGRM were 0.910 ± 0.447, 1.004 ± 0.418 and 0.987 ± 0.370, respectively. The correlations and regression slopes by traits with sufficient data to make inference are presented in Table 1. These results indicate that the BGRM resulted in slightly higher cross-validation accuracies at the expense of a negligible increase in bias over the SGRM EBVs. A stronger bias was found in the NRM EBVs than for either of the single-step analyses. While the standard deviations were large across traits for both accuracies and biases, together these set of results suggest that correcting the relationship values of the GRM for breed can produce higher accuracy and lower bias in EBVs.

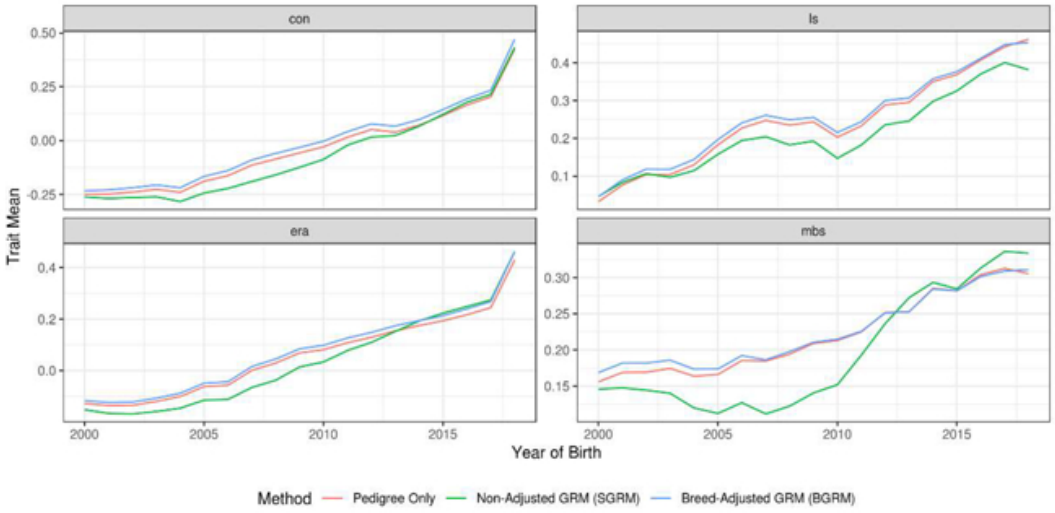

Figure 1. Genetic trends for the four traits showing the lowest correlation between EBVs when using a single breed GRM. EBVs have been scaled by dividing by the genetic standard deviation

The implementation of the method presented by Makgahlela *et al.* (2013) for a breed-adjusted single-step genomic evaluation has some advantages compared to an approach using metafounders. Firstly, the adjustment of the GRM is a simpler modification to the single-step relationship matrix than that required by metafounders and allows any current genetic grouping structure (pedigree-derived genetic groups for example) to exist alongside the modified relationship matrix. Adjusting the GRM is also simpler when breeds or genetic groups have no genotyped animals in their pedigree. Implementing a BGRM in a single-step analysis requires that genetic groups are also fitted in the model as they have for NRM based analyses. These genetic groups need to align with the breeds that were used in the construction of the GRM. There are situations where the implicit breed structure in the GRM has advantages, e.g. predicting breeding values for animals without pedigree across genetic groups, with metafounders allowing this structure to be imposed over the whole NRM.

Genetic groups or metafounders both require the assignment of animals into pre-defined group structures. Methods for creating the most parsimonious grouping structures require further

investigation, minimising the number of groups required while maintaining enough groups for predictive purposes. The addition of genotypes can aid in this process.

CONCLUSIONS

In this paper, we show that the method presented by Makgahlela *et al.* (2013) reduces the breed structure implicit in a GRM constructed from multiple breeds, resulting in a GRM that is numerically more similar to the NRM. This change results in genetic trends that align closer with those seen from pedigree-only models. The BGRM resulted in slightly higher average cross-validation accuracies with similar biases, and less biased than pedigree alone, compared to BLUPs performed using a GRM constructed from a single set of allele frequencies.

Table 1. Table of forward cross-validation accuracies obtained from BLUP models using an NRM (r_NRM), a single-breed GRM (r_SGRM) and a multi-breed GRM (r_BGRM) and the corresponding biases, b_NRM, b_SGRM and b_BGRM. 'n' indicates the number of animals in the validation set

Trait	n	r_NRM	r_SGRM	r_BGRM	b_NRM	b_SGRM	b_BGRM
ycon	618	0.10	0.14	0.16	0.59	0.83	0.95
con	885	0.11	0.12	0.18	0.53	0.62	0.87
yls	627	0.18	0.20	0.17	0.79	0.84	0.78
ls	1,801	0.14	0.14	0.16	0.69	0.74	0.79
yera	377	0.22	0.21	0.25	1.84	2.05	2.05
era	1,583	0.34	0.27	0.29	2.00	1.90	1.72
pemd	3,476	0.15	0.17	0.18	0.74	0.83	0.79
pcf	3,467	0.21	0.24	0.25	0.94	1.13	1.10
pwt	431	0.25	0.27	0.26	0.80	0.86	0.83
awt	943	0.32	0.34	0.35	0.71	0.81	0.77
cs	545	0.34	0.34	0.34	1.08	1.20	1.16

Abbreviations: 'ycon' and 'con': fertility of yearling and adult ewes, respectively, 'yls' and 'ls': little size of yearling and adult ewes, respectively, 'yera' and 'era': rearing ability of yearling and adult ewes, respectively, 'pwt' and 'awt': pre-joining weight of post-weaning and adult ewes, respectively, 'cs': pre-joining condition score of adult ewes.

REFERENCES

Boerner V., Wittenburg D. (2018) *Front. Genet.* **9**:185.

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

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., Cantet R.J.C. (2017) *Genet. Sel. Evol.* **49**:34.

Legarra A., Christensen O.F., Aguilar I., Misztal I. (2014) *Livest. Sci.* **166**:54.

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

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

Meyer K., Tier B., Swan A. (2018) *Genet. Sel. Evol.* **50**:39.

Vitezica Z.G., Aguilar I., Misztal I., Legarra A. (2011) *Genet. Res.* **93**:357.

Westell R.A., Quaas R.L., Van Vleck L.D. (1988) *J. Dairy Sci.* **71**:1310.

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.