Proc. Assoc. Advmt. Anim. Breed. Genet. 25: 334 - 337

UTILITY OF POOLED DNA SAMPLES FOR ESTIMATING A FLOCK PROFILE

P.M. Gurman, K. Gore and D.J. Brown

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

SUMMARY

The flock profile product by Sheep Genetics allows commercial Merino breeders to benchmark their flock's genetic merit based on the genotypes of 20 animals. Sheep breeders collect DNA samples from their sheep using Tissue Sampling Units, which are then sent to the DNA laboratory and converted into genotypes for the 20 animals, which are used to calculate individual animal breeding values. The final reported value provided to the breeder is the mean of the estimated ASBVs for the 20 animals. This study documents an in-silico investigation to determine if the individual animal genotypes can be combined into an allele frequency, which is used instead to estimate the flock profile breeding value. The mean correlation across traits was 0.99999, while the mean regression slope was 0.9999 These results show that it is possible to calculate the flock profile breeding values based on the allele frequencies. Further research is now required to research and develop procedures on a commercial scale and examine the correlation between a genotype from a pooled sample and the allele frequencies calculated from individual genotypes at this scale.

INTRODUCTION

The flock profile test is a genomic test offered to Australian Merino sheep breeders, which provides a benchmark of their flock's genetic merit compared to the MERINOSELECT analysis (Swan *et al.* 2018). This product requires that DNA samples are collected using Tissue Sampling Units (TSU) on 20 randomly selected sheep from the most recent drop, which are then sent to a genotyping laboratory and analysed as 20 individual animals. The resulting genotypes are then used to calculate Australian Sheep Breeding Values (ASBVs) for each animal based on the reference population of genotyped and phenotyped animals from the MERINOSELECT single-step analysis (Swan *et al.* 2018), assuming unknown pedigree. The ASBVs for the individual animals are then averaged to estimate the flock profile. This process results in ASBVs that are directly comparable to ASBVs reported in the full MERINOSELECT single-step analysis and validated by leaving the data of one flock out of the analysis at a time and estimating breeding values from the remaining data (Swan *et al.* 2018). This service has been used since its inception in 2016 for over 600 commercial flocks.

Currently, the cost of a flock profile includes the cost of genotyping 20 animals. One option for reducing the cost of this product and increasing its adoption is to pool the DNA from the 20 animals. The pooled sample can then be processed by the genotyping laboratory to obtain the dosage/allele frequency based on these 20 animals. For this to be a viable option, the ASBV estimated from a pooled sample needs to be equivalent to the ASBV calculated from the mean of the 20 animals calculated separately. This study examines if the mean of the 20 animal's ASBVs as is currently done to calculate a flock profile is sufficiently like the ASBV calculated from the mean of the individual genotypes (allele frequency) from the 20 animals, which would be available from a single genotype from a pooled DNA sample in practice.

MATERIALS AND METHODS

In this study, previous flock profile tests (n flocks = 673, n animals = 13,017) were used, extracting the genotypes for each individual animal from the MERINOSELECT analysis. These

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

genotypes have previously been cleaned (individual call rate 90+%, heterozygosity <=50%) and imputed to fill in sporadic missing SNP calls using all available genotypes for the chip on which the animal was genotyped. The genotypes were then imputed to the 4 other separate SNP chips that have significant reference populations (n>10000 Australian and New Zealand genotypes for each reference set). The separate imputation results were then combined into a set of 60,410 SNP genotypes, starting from the original genotype and adding the SNPs from the other chips that were not already present. All imputation was performed using Beagle (Browning *et al.* 2018; 2021). Genotypes for the animals included in the MERINOSELECT analysis were then used to calculate SNP effects based on their ASBVs. The reference population for the traits analysed ranged from 11,192 to 143,356 genotyped and phenotyped animals with a mean of 74,338 animals. Genotypes for each flock profile were then used to calculate the mean of the genotypes for each flock profile, i.e. twice the allele frequency for each flock profile, and the resulting genotype values as double precision floating point values between 0 and 2 were used to calculate an ASBV based on Swan *et al.* (2018). These new pooled results were then compared to the traditional method as part of the current MERINOSELECT analysis. Analyses were performed for all traits which are reported for flock profile tests and traits used in current selection indexes, (body weight at weaning, post-weaning, yearling and adult age stages; greasy fleece weight at adult and yearling, clean fleece weight at yearling; fibre diameter, its coefficient of variation, staple length and curvature at yearling and adult; carcase fat and eye muscle depth at post-weaning and yearling; early breed cover and breech wrinkle and late body wrinkle. Metrics examined between the two sets of ASBVs included Pearson correlations, dispersions calculated as $cov(\boldsymbol{u}_{current}, \boldsymbol{u}_{pooled})/var(\boldsymbol{u}_{pooled})$ and the scaled bias as $\frac{\overline{\boldsymbol{u}_{pooled}}-\overline{\boldsymbol{u}_{current}}}{\sigma_g}$. Data preparation, calculation of EBVs and statistical analysis of the results was performed using Python 3.10 and the Pandas library 1.5.0.

RESULTS AND DISCUSSION

The mean correlation across traits between the current flock profile ASBVs and those obtained from allele frequencies was $0.999985 \pm 6.17 \times 10^{-5}$, with these correlations presented in Figure 1. The outlier trait was post-weaning faecal egg count (PWEC), which had a correlation of 0.9997. The mean dispersion was 0.9999 ± 0.003. For most traits, there was a slight increase in the dispersion of the ASBVs estimated, with the dispersions presented in Figure 2. PWEC was again the outlier with a lower variation in the ASBVs estimated from the allele frequency. Finally, the mean scaled bias was -0.0180 ± 0.108, though this deviation from zero was largely driven by the PWEC bias value (-0.67). The scaled biases are presented in Figure 3. These results show little difference between the ASBVs estimated from the mean of the ASBVs from individual animals and those estimated from the allele frequencies. This is not surprising as the calculation of breeding values is a linear function of the SNP effects. The reason for the slightly reduced precision in PWEC is likely, in part, due to the non-normality of the phenotypic distribution of PWEC data. While the transformation of the data used, largely addresses this problem, the slightly lower precision is unlikely to have a realised effect on selection decisions on farm.

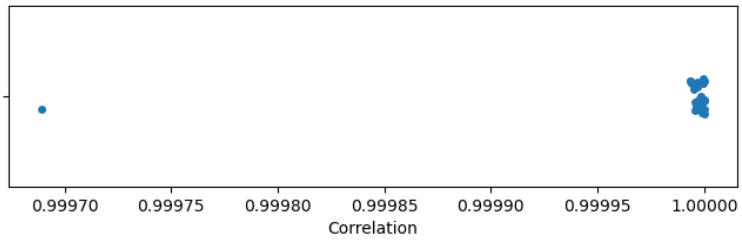

Figure 1. Correlations between ASBVs obtained from allele frequencies and from the mean of the individual animal genotypes for all flock profile traits

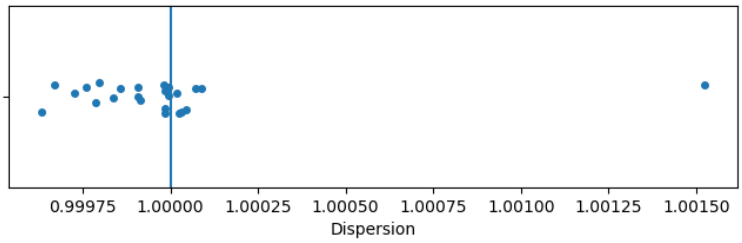

Figure 2. Dispersions between ASBVs obtained from allele frequencies and from the mean of the individual animal genotypes for all flock profile traits

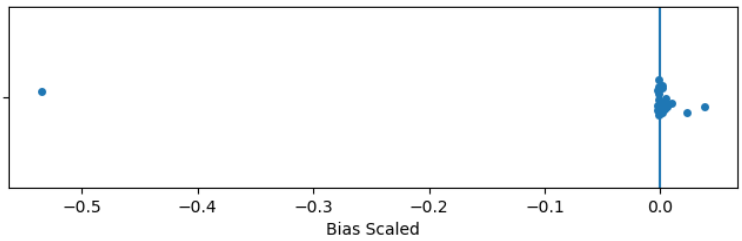

Figure 3. Biases scaled by the genetic standard deviation between ASBVs obtained from allele frequencies and from the mean of the individual animal genotypes for all flock profile traits

Various potential issues arise in processing the DNA for a pooled sample. One issue is having all individuals being represented equally within the pooled sample. High volume genotyping labs don't normalize the concentration of DNA when processing individuals (Neogen Australia, pers comms). The additional cost of normalization of DNA concentrations before pooling would mean that a direct 95% reduction in price of the flock profile would not be feasible. We expect that the cost reduction, would still be at a point where it would be beneficial, as other uses for DNA pooling have demonstrated (Bell *et al.* 2017; Aldridge *et al.* 2022). This could also allow for a larger proportion of the flock to be included, rather than the current 20 individuals, which would potentially be a better representation of that flock.

While in this paper we have used the mean of the genotype, extracting the frequency from the data generated by the genotyping platform may not be as straight forward. Janicki *et al.* (2008) present multiple methods for extracting or calculating the SNP allele frequencies. One method indicates that the Illumina Genome Studio Genotyping Module (Illumina Inc) automatically produces the B allele frequency in its reporting which was demonstrated to be acceptable as the

frequency. Alternative platforms (e.g. Affymetrix or genotyping by sequence) would need to explore alternative methods.

Imputation of the pooled genotype is another issue. The current version of the Beagle software requires genotypes as input, coded 0,1,2. Version 4 of Beagle is capable of accepting a genomic likelihood, which may be usable for imputing the pooled genotype, and providing a genomic probability (Browning *et al.* 2016). Wen *et al.* (2010) have also presented algorithms specifically for pooled genotype data.

This paper demonstrates that flock profile ASBVs may be able to be calculated from a pooled genotype, however validation of the pooled sample methods would require individual and pooled genotyping results. The most cost-effective way of achieving this would be to resample existing flock profile animals using the pooling process, or to attempt this new method on breeder submitted flock profile tests alongside the current individual animal genotyping process.

CONCLUSION

This research suggests that collapsing genotypes down to the mean of the genotypes has little impact on the ASBV calculated for a flock profile. Further research is needed to determine if the pooling of DNA samples before genotype estimation can be used to reduce the costs of calculating a flock profile, including challenges of application in a commercial environment.

ACKNOWLEDGEMENTS

Funding for this research was provided by Meat and Livestock Australia under L.GEN 2204 and L.GEN 2203.

REFERENCES

Aldridge M.N., Marjanovic J., Henshall J.M., de Klerk B., Peeters K. and de Haas Y. (2022) *Proc. 12th World Congr. Genet. Appl. Livest. Prod*. **12:** 781.

Bell A.M., Henshall J.M., Portro-Netro L.R., Dominik S., McCulloch R., Kijas J. and Lehnert S.A. (2017) *Genet. Sel. Evol.* **49:** 1.

Browning B.L. and Browning S.R. (2016) *Am. J. Hum. Genet.* **98:** 116.

Browning B.L. Tian X., Zhou Y. and Browning S.R. (2021) *Am. J. Hum. Genet.* **108:** 1880.

Browning B.L., Zhou Y. and Browning S.R. (2018) *Am. J. Hum. Genet.* **103:** 338.

Janicki P.K. and Liu J. (2008) *Internet J. of Genom. and Proteomics* **5:** 1.

Swan A.A., Gurman P.M., Boerner V., Brown D.J., Clarke S., Gore K., Granleese T. and van der Werf J.H.J. (2018) *Proc. 11th World Congr. Genet. Appl. Livest. Prod*. **11:** 346.

Wen X. and Stephens M. (2010) *Ann. Appl. Stat.* **4:** 1158.