Proc. Assoc. Advmt. Anim. Breed. Genet. 26: 455-458

EXPANDING AUSTRALIAN SHEEP GENOMIC REFERENCE POPULATION

P. Alexandri, P.M. Gurman, D.J. Brown, K.L. Bunter, K.P. Gore, S. de las Heras-Saldana, 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

Allele frequencies play a crucial role in multi-breed genetic evaluations by enabling the construction of a breed-adjusted genomic relationship matrix (GRM), which can be derived from a genomic reference population. This study aimed to update the Australian sheep genomic reference population for a combined Terminal and Maternal LAMBPLAN genetic evaluation. The analysis incorporated genotypic data from sheep across maternal, terminal, shedding, and Merino breeds, as well as smaller breeds and composite populations. Population structure analysis identified 27 genetic clusters among terminal, maternal and shedding breeds and five clusters within Merino sheep. The findings demonstrate that genomic data and unsupervised clustering methods can improve breed composition accuracy and facilitate a comprehensive multi-breed genetic evaluation, reducing the need for strictly straight-bred reference populations.

INTRODUCTION

Genomic information has increased the accuracy of Australian Sheep Breeding Values (Brown *et al.* 2018). Genetic evaluation for the large and diverse Australian sheep population (Swan *et al.* 2015) requires individual sheep to be assigned breed proportions based on a genomic reference population used to identify breed content, for a breed adjusted GRM with a reduced breed structure to be constructed (Gurman *et al.* 2019). The Australian sheep genomic reference population included pedigree defined Border Leicester, Poll Dorset, Texel, Suffolk, White Suffolk and Merino animals (Gurman *et al.* 2017). Since then, the number of genotypes available for the LAMBPLAN genetic evaluation has vastly increased (Walkom *et al*. 2025). These genotypes come from both straight-bred individuals belonging to the above major sheep breeds, but also from smaller breeds and composite populations whose numbers in the genetic evaluations have been steadily increasing in the past few years (McMillan *et al.* 2023).

This work aimed to redefine the Australian sheep genomic reference population to be used for a breed adjusted GRM in preparation for a combined Terminal and Maternal LAMBPLAN genetic evaluation (Walkom *et al*. 2025) using genotype data from maternal, terminal, shedding and Merino sheep breeds, as well as from smaller breeds and composite populations. The latter were until recently, not included in the LAMPBPLAN single step genetic evaluation because reference populations for inclusion in breed adjusted relationship matrices had not been established. Two different datasets were used in a population genetic analysis to investigate population structure across different breeds and to calculate allele frequencies to define breed composition.

MATERIALS AND METHODS

Genotypes used in this analysis came from 260,200 sheep of different terminal and maternal breeds. A GRM was constructed within breed based on the method implemented by Yang *et al.* (2011). To avoid bias in the analysis that might be introduced by family structure, individual genotypes were filtered to represent different breeds but to be as distantly related as possible. Within breed, animals were retained using the following criteria: i) animals with no parents present in the

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

pedigree or no parents genotyped, ii) animals with an average pedigree Q matrix value > 0.9, and iii) average genomic relationship lower than 0.2. Because animals included in the data were genotyped with different single nucleotide polymorphism (SNP) panels, genotypes were filtered to include only the common SNPs (2,443) between these panels, resulting in a data set of 3,777 sheep from Terminal, Maternal and smaller breeds (Dataset 1, original genotypes). To account for Merino crosses in the LAMBPLAN analysis and to identify possible population structure within the Merino breed, a second data set of 20,501 unrelated Merino animals, genotyped in the same SNPs, was used (Dataset 2, original genotypes). To facilitate calculation of allele frequencies for all SNP data used in regular LAMBPLAN evaluations, all genotyped animals were imputed to 62,016 SNPs (imputed genotypes). The numbers of genotypes from each breed are presented in Table 1.

Table 1. Breeds, breed types and codes and number of unrelated animals (N) from each breed used in the analysis

Breed Code	Breed Name	Breed Type	N	Dataset
02	Border Leicester	Maternal	558	
03	Corriedale	Wool	26	
09	Wiltshire Horn	Shedding	32	
10	Polwarth	Wool	25	
11	Hampshire Down	Terminal	72	
12	Wiltipoll	Shedding	122	
14	Southdown	Terminal	77	
15	Coopworth	Maternal	72	
16	Poll Dorset	Terminal	896	
17	Texel	Terminal	59	
18	Romney	Wool	9	
19	Suffolk	Terminal	147	
22	Finnsheep	Maternal	20	
23	White Suffolk	Terminal	284	
24	Ultra White	Shedding	354	1
25	Australian White	Shedding	181	
26	Research	Composite	59	
31	English Leicester	Maternal	17	
38	East Friesian	Maternal	98	
40	Dorper	Shedding	132	
47	White Dorper	Shedding	147	
48	SAMM	Wool	65	
49	Damara	Shedding	29	
50	Merino	Wool	8,130	
51	Dohne Merino	Wool	19	
CM	Commercial Maternal	Maternal	25	
CS	Commercial Shedders	Shedding	43	
CT	Commercial Terminal	Terminal	145	
IF	Ile De France	Terminal	27	
VR	Van Rooy	Shedding	17	
60	Poll Merino	Wool	11,974	
74	NZ Merino	Wool	308	2
76	NZ Lincoln	Terminal	20	
85	NZ Poll Merino	Wool	89	
Total			24,278	

Population structure for Dataset 1 and 2 was inferred using the method implemented in STRUCTURE 2.3.4 software (Pritchard *et al*. 2000). STRUCTURE clusters individuals in groups minimising both linkage and Hardy-Weinberg disequilibrium (Falush *et al.* 2003) utilising a Bayesian approach. Individuals were assigned to a pre-determined number of clusters (K). For each cluster, allele frequencies were estimated, and individuals were re-assigned based on those frequency estimates until convergence was achieved. The existence of possible clusters within the Australian sheep population, was estimated using K values up to 40. Any previous breed information was discarded (unsupervised analysis), and the admixture model was used with correlated allele frequencies and eight independent runs of 10,000 iterations, after a burn-in of 4,000 iterations for each K. The optimal K value was informed based on the approach of Evanno *et al.* 2005. Final allele frequencies for the optimal K were estimated by aligning the replicates of each run and examining pairwise similarities between allele frequency matrices with the method described by Jakobsson and Rosenberg (2007). To avoid bias due to overrepresentation, for each of the clusters identified for Dataset 2, individuals were selected if their membership coefficient was greater than 90%. Since STRUCTURE analysis was performed for both the original and the imputed genotypes (results not shown), the consistency of estimated clusters was evaluated by calculating the correlation of the cluster membership Q matrices for each breed between the original and imputed data.

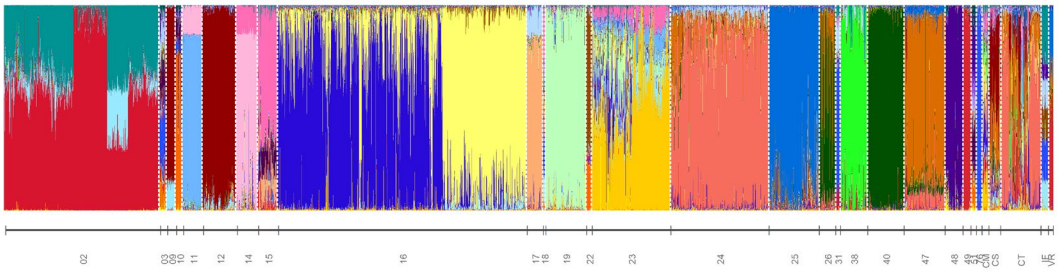

Figure 1. STRUCTURE analysis results for Dataset 1 with optimal number of clusters K=27. Each vertical line represents one individual and colours correspond to each of the identified clusters. Breed codes correspond to Table 1

RESULTS AND DISCUSSION

Analysis of the combined Terminal and Maternal breed genotypes (Dataset 1, original genotypes) defined 27 clusters (K) (Figure 1). The results showed several relatively homogeneous clusters, in some cases corresponding to different breeds like East Friesian (breed Code: 38), Hampshire Down (11), Southdown (14), Texel (17), Suffolk (19), Wiltipoll and Wiltshire Horn (12 and 09 respectively), Dorper (40) and Australian White (25). Several breeds were shown to include more than one cluster like Ultra White (24) which appear to share ancestry with White Dorper (47). This was also observed within the Commercial Maternal, Commercial Terminal and Commercial Shedder groups which contain different clusters corresponding to their composite origin. Analysis of Dataset 2 revealed the existence of five clusters within the Merino animals that seem to be more related to breeding objectives than geographical origin (Figure 2). This method identified 756 Merino animals from all five clusters to be included in the genomic reference population. Results were compared between original and imputed data for the same number of clusters, and mean cluster membership correlations across breeds was 0.93 (± 0.06) ranging from 0.77 to 1.00 for animals of different breeds. These high correlations demonstrate that similar results can be obtained when imputed data is used, and allele frequencies for each cluster calculated with this data correspond to those calculated with the original genotypes. Allele frequencies for the 62,016 SNPs in the imputed genotype data set were estimated for each of the clusters identified in the STRUCTURE analysis

and used to calculate breed proportions for all genotyped animals used in the LAMBPLAN genetic evaluation.

The use of genomic data to identify genetic structure in livestock breeds had been previously used in beef and sheep (Sölkner *et al.* 2010) and have been used to assess breed diversity, determine breed composition and improve genetic evaluations (Gurman *et al.* 2019). A thoroughly selected set of genotyped animals combined with an unsupervised admixture approach can eliminate the need for straight-bred animals. In this light, smaller breeds and composite populations can be better accounted for and facilitate a combine Terminal and Maternal multibreed genetic evaluation.

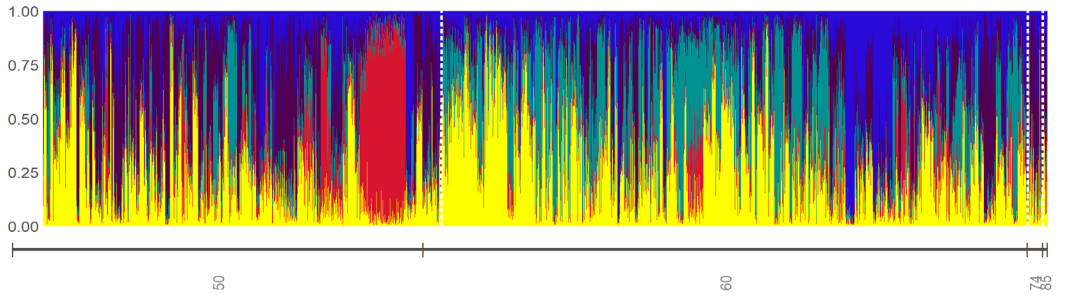

Figure 2. STRUCTURE analysis results for the Merino dataset with optimal number of clusters K=5. Each vertical line represents one individual and colours correspond to each of the identified clusters. Breed codes correspond to Table 1

CONCLUSION

The genomic reference population for Australian sheep has been successfully updated by incorporating genotypic data from maternal, terminal, shedding, and Merino breeds, and including smaller breeds and composite populations. By leveraging genomic data and unsupervised clustering methods, the accuracy of genomic breed composition assessment can be enhanced resulting in a more inclusive multi-breed genetic evaluation. The findings support a shift away from relying solely on straight-bred reference populations, allowing for better representation of composite and smaller breeds in genetic evaluations.

ACKNOWLEDGEMENTS

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

REFERENCES

Brown D.J., Swan A.A., Boerner V., *et al.* (2018) *Proc. World Cong. Genet. Appl. Livest. Prod.* **11**: 16.
Evanno G., Regnaut S. and Goudet J. (2005) *Mol. Ecol* **14**: 2611.
Falush D., Stephens M. and Pritchard J.K. (2003) *Genetics* **164**: 1567.
Gurman P.M, Swan A.A. and Boerner V. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22**: 341.
Gurman P.M., Bunter K.L., Boerner V., *et al.* (2019) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **23**: 254.
McMillan A.J., Walkom S.F. and Brown D.J. (2023) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **25**: 31.
Jakobsson M. and Rosenberg N.A. (2007) *Bioinformatics* **23**: 1801.
Pritchard J.K., Stephens M. and Donnelly P. (2000) *Genetics*, **155**: 945.
Sölkner J, Frkonja A, Raadsma H.W, *et al*. (2010) Interbull Bulletin. **42**: 62.
Walkom S.F., Alexandri P., *et al.* (2025) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **25**: *These proceedings*.
Yang J., Lee S. H., Goddard M. E. and Visscher P.M. (2011). *Am. J. Hum. Genet*. **88**: 76.