Proc. Assoc. Advmt. Anim. Breed. Genet. 20:66-69

STRATEGIES TO OBJECTIVELY GROUP MERINO FLOCKS IN SHEEP GENETICS

D.J. Brown[1], A.A. Swan[1], J.S. Gill[2] and R.B. Banks[1]

[1]Animal Genetics and Breeding Unit[*], University of New England, Armidale, NSW 2351
[2]Meat and Livestock Australia, University of New England, Armidale, NSW 2351

SUMMARY

The Merino breeding population in Australia exhibits considerable diversity in objectives and breeding philosophies. It could be beneficial for both the analysis and reporting of the national genetic evaluation to objectively group flocks into logical subsets. This study evaluates techniques to cluster flocks into logical groups based on either estimated breeding values or genomic information. Principle component analyses were conducted using flock mean breeding values and the genomic relationship matrix. Using the flock mean breeding values, 6 clusters of flocks were identified with the first 2 principle components explaining 73% of the variation between flocks. The first principle component separated flocks based on overall productivity, with approximately equal emphasis across all traits. The second component separated flocks based on fleece weight, wrinkle and staple length. Less separation between flocks was apparent for flocks with below average fibre diameter. The principle components of the genomic relationship matrix were also strongly correlated with mean breeding values across the flocks. The lack of accurate Australian Sheep Breeding Values (ASBVs) for some traits and genomic information across some of the flocks is a limitation of this approach as it makes allocation of some flocks challenging.

INTRODUCTION

Over recent decades mixing of animals between strains of Merino sheep has become more widespread. The industry is made up of flocks with a range of breeding objectives and many breeding philosophies with varying levels of objective measurement and visual selection in use. Sheep Genetics is a genetic evaluation service which provides Australian Sheep Breeding Values (ASBVs) to sheep breeders (Brown *et al.* 2007). Sheep Genetics uses a "Type" classifier to allow separation of Merino flocks for reporting and comparison of genetic trends. However, at times this classifier is too simple to adequately group flocks. This type classification is mostly based on ASBVs ranges for key traits and breeder perception of their wool and sheep type.

The aim of this study was to evaluate techniques to objectively group flocks based on either their average breeding values for key traits or genomic information.

MATERIALS AND METHODS

Estimated breeding values. ASBVs were extracted from the Sheep Genetics MERINOSELECT database (Brown *et al.* 2007). This database consists of pedigree and performance records submitted by Australian ram breeders which are used for genetic evaluation purposes. Traits extracted were Yearling live weight (Ywt), clean fleece weight (Ycfw), fibre diameter (Yfd), fibre curvature (Ycuv), staple length (Ysl) and breech wrinkle (Bwr). For each flock the average ASBV for each trait was calculated using all animals born since 2006. A summary of the data used for each key trait is shown in Table 1. There were 256 flocks with ASBVs available for all traits available.

[*] AGBU is a joint venture of NSW Department of Primary Industries and University of New England

Table 1. Summary of the variation in flock mean ASBVs for the 256 flocks used in this study

Trait	Mean	SD	Min	Max
Ywt (kg)	1.75	2.59	-4.81	9.40
Ycfw (%)	5.48	7.92	-17.53	25.85
Yfd (µm)	-1.05	0.68	-3.14	1.30
Ycuv (deg/mm)	-0.76	4.67	-11.58	17.34
Ysl (mm)	2.15	4.21	-12.52	16.71
Bwr (scores)	-0.05	0.29	-1.13	1.04

Genomic information. 50K Ovine SNP chip (Illumina Inc., San Diego, CA, USA) genotypes were available on 6230 merino animals from 203 flocks. Most of the animals genotyped were part of a reference population in which sires from these flocks were mated to ewes representative of several industry types. While 137 flocks had both ASBVs and at least one animal genotyped, only 51 flocks had ASBVs and 5 or more genotyped animals. In the main however, the set of animals included in ASBV means were independent from the set of animals genotyped.

Models of analysis. Principle component analysis of the standardised flock means for all six traits was conducted using the princomp procedure in R (R Development Core Team 2012) with the resultant 6 principle components (PC1 to PC6) used in the kmeans procedure to cluster flocks. As there are significant differences between traits in their means and variance, all traits were standardised (mean = 0, standard deviation = 1) to allow equal contribution of all traits to the principle component analysis. There is normally a reduction of the within-cluster variation as more clusters are used. However a balance is required between this and the number of flocks within each cluster and the interoperability of the clusters formed. After examining the results from using 4 to 7 clusters, 6 was chosen as the optimal number to adequately separate flocks into groups.

The genomic relationship matrix was calculated for all animals in the Sheep Genetics database with 50K SNP genotypes following the methods of Van Raden (2008) and Yang et al. (2010) and scaled so that the average diagonal element was 1. Principle components of this relationship matrix were then estimated using the singular value decomposition method in a purpose written program using the LAPACK numerical computation libraries. This program was used rather than R for speed of computation.

The first five principle components of the genomic relationship matrix were used for this study. The principle components were standardised prior to being averaged over flocks with animals represented. Flocks with fewer than 2 animals were removed leaving 87 flocks with genomic and ASBV based principle components.

RESULTS AND DISCUSSION

ASBV based principle components 1 to 6 explained 58%, 16%, 10%, 8%, 6% and 2% of variation respectively. The first 2 principle components explain most of the variation (74%). Table 2 shows the correlations between traits and principle components. The flock means were moderately correlated across flocks with most traits except wrinkle having correlations greater than 0.5 with the other traits. PC1 was highly correlated with all ASBVs and thus represented a discriminator of overall production across the Merino industry. The second principle component was more related to wrinkle, staple length and fleece weight, and a little with curvature and staple length, and is likely to separate flocks on style and breeding philosophy. The remaining principle components appeared to concentrate on individual traits.

Table 2 Correlations between traits (standardised) and with principle components

	Ywt	Ycfw	Yfd	Ycuv	Ysl	Bwr
Ywt		0.62	0.60	-0.46	0.54	-0.32
Ycfw	0.62		0.50	-0.72	0.47	-0.11
Yfd	0.60	0.50		-0.63	0.56	-0.25
Ycuv	-0.46	-0.72	-0.63		-0.74	0.28
Ysl	0.54	0.47	0.56	-0.74		-0.42
Bwr	-0.32	-0.11	-0.25	0.28	-0.42	
PC1	-0.78	-0.78	-0.80	0.87	-0.83	0.45
PC2	-0.02	-0.40	-0.08	0.14	0.18	-0.85
PC3	-0.58	-0.08	-0.12	-0.39	0.29	0.05
PC4	-0.02	-0.40	0.54	0.08	0.08	0.17
PC5	0.18	-0.19	-0.25	0.11	0.41	0.21
PC6	-0.09	0.15	0.06	0.25	0.11	-0.02

Cluster means for each trait are generally the best way to visualise the characteristics of the clusters formed. Table 3 shows these cluster means for each trait. For example cluster 3 appears to be the traditional superfine flocks which are low for live weight, fleece weight, fibre diameter and staple length but higher for curvature and wrinkle.

Table 3. Cluster means for each trait (standardised)

Cluster	Flocks	Ywt	Ycfw	Yfd	Ycuv	Ysl	Bwr
1	31	0.61	0.31	1.00	-1.17	1.56	-0.66
2	7	-0.25	-0.35	0.09	0.14	-0.13	-0.06
3	35	-1.22	-1.33	-1.34	1.39	-1.17	0.58
4	20	-0.38	0.44	-0.21	-0.18	-0.48	1.13
5	30	0.85	-0.13	0.15	0.47	-0.13	-1.67
6	14	0.72	1.01	0.40	-0.59	0.37	0.03

The principle components of the genomic relationship matrix were also moderately to highly correlated with the mean ASBVs for the flocks (Table 4). In particular the first genomic principle component appears to separate flocks based on overall production level. The second component appears to separate on fleece weight, curvature, staple length and wrinkle which may relate to flocks using traditional versus skin related breeding philosophies. This is consistent with the clustering based on flock mean breeding values.

Table 4. Correlations between the ASBV flock means and averaged principle components from the genomic relationship matrix (n=137)

	Ywt	Ycfw	Yfd	Ycuv	Ysl	Bwr
genPC1	-0.69	-0.69	-0.70	0.77	-0.65	0.36
genPC2	-0.03	0.21	0.01	-0.34	0.16	-0.18
genPC3	-0.15	0.00	0.12	-0.14	-0.13	-0.01
genPC4	0.11	-0.04	0.04	-0.15	0.36	-0.28
genPC5	0.18	0.00	0.04	0.06	-0.02	0.02

Principle component analysis relies on having all traits observed which reduces the number of flocks which we can cluster. There is also significant variability in the number of animals recorded

and the ASBV accuracy for some traits such as staple length and breech wrinkle. Furthermore the number of animals genotyped across the flocks varies greatly thus influencing the accuracy of the genomic relationship between flocks.

CONCLUSIONS

The results from these analyses clearly show that the Merino industry no longer has distinct types and for all traits considered there is a continuum of performance. Despite this the analysis was able to separate the flocks into 6 distinct groups. While clusters may have been similar for some components they were distinctly divergent for at least one principle component. Clustering based on both flock mean breeding values and on genomic relationships was consistent, with two main clustering dimensions emerging as being the key production traits and plainness of body. This is not altogether surprising since the genetic analysis upon which the breeding value clustering is based, includes genetic relationships. The consistency suggests a reasonably high level of pedigree accuracy in the Sheep Genetics data.

The main reason for exploring this approach was to consider how best to group flocks for both analysis and reporting of results. Sensible clusters would help breeders interpret and filter the mass of breeding value information to make appropriate selection decisions. Both the ASBVs and genomic-based clusters appear intuitively sensible. Further work and consultation with breeders is warranted to determine how best to combine the two sources of clustering information, how to use the resulting clusters in analysis, and whether this approach offers an improvement in clarity for breeders.

ACKNOWLEDGMENTS

This research was funded by Meat and Livestock Australia and Australian Wool Innovation, and Sheep Genetics is made possible through the support of the Australian sheep industry. The authors acknowledge the contributions of the Sheep CRC Information Nucleus, the Australia Merino Sire Evaluation Association and industry-funded research flocks.

REFERENCES

Brown D.J., Huisman A.E., Swan A.A., Graser H-U., Woolaston R.R., Ball A.J., Atkins K.D. and Banks R.B. (2007) *Proc. Ass. Adv. Anim. Breed. Genet.* **17**: 187.

R Development Core Team (2012) Statistical Computing, http://www.R-project.org.

VanRaden P.M. (2008) *J. Dairy Sci.* **91**: 4414.

Yang J., Beben B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.F., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E. and Visscher P.M. (2010) *Nature. Genetics.*, **42:** 565.