

Single-Step Genetic Evaluations in the Australian Sheep Industry

D.J. Brown^{1,2}, A.A. Swan^{1,2}, V. Boerner¹, L. Li^{1,2}, P.M. Gurman^{1,2}, A.J. McMillan^{1,2} and J.H.J. van der Werf^{2,3}, H.R. Chandler⁴, B. Tier¹ & R.G. Banks¹

¹Animal Genetics and Breeding Unit, University of New England, Armidale 2351, Australia
dbrown2@une.edu.au (Corresponding Author)

²Cooperative Research Centre for Sheep Industry Innovation, Armidale 2351, Australia

³School of Environmental and Rural Science, University of New England, 2351 Armidale, Australia.

⁴Meat and Livestock Australia, University of New England, Armidale 2351, Australia

Summary

The main benefit of genomic selection for Australian sheep is to increase the accuracy of estimates of genetic merit for hard to measure traits including carcass, adult wool and reproduction traits. An extensive genotyped reference population, combined with a significant number of genotyped and phenotyped animals from ram breeding flocks enables genomic predictions with improved accuracies. Genomically enhanced breeding values have now been transitioned from a mixture of blending and independent single-trait single-step methods into full multiple-trait single-step analyses covering most traits evaluated. Key challenges to achieve this were: weighting of pedigree and genomic information, achieving acceptable run times, and estimation of breeding value accuracy from genomic contributions. The new analyses have been shown to significantly improve the prediction of progeny performance across most traits. Future developments will evaluate alternate models for incorporating genomic information, ability to include genotyped animals without any pedigree and phenotypic information, and development of single-step analyses for reproduction traits.

Keywords: genomics, ebv, single-step

Introduction

The Australian sheep industry has a well-developed genetic evaluation system to estimate Australian Sheep Breeding Values (ASBVs) (Brown *et al.*, 2007), managed by Meat and Livestock Australia's Sheep Genetics business unit, and features three separate large scale analyses: Merinos, maternal breeds (with the major breeds being Border Leicester, Coopworth, and more recently maternal composites), as well as terminal breeds (dominated by the Poll Dorset and White Suffolk breeds). The three analyses are all conducted across breeds and flocks, and there are currently 2.5 million animals in the pedigree for Merinos, 2.1 million for maternal sires, and 3.1 million for terminal sires, with approximately 157,000, 73,000, and 127,000 new animals entering the system each year, respectively. For each breed 4 separate analyses are conducted for different trait groups (production, worm egg count, visual and reproduction trait groups).

Significant genetic gains have been made, particularly in the terminal and maternal sire sectors (Swan *et al.*, 2017), and from 2007, the industry has invested heavily in research to develop genomic selection to further increase rates of gain. Unlike in dairy cattle, where genomic selection has a major impact through shortening of generation intervals because it is

possible to have accurate evaluations of bulls with reduced progeny testing (Schaeffer, 2006), sheep are typically evaluated with reasonable accuracy from an early age in both sexes, at least for traits which are easy to measure. This means that the added gains from genomic selection in sheep are more modest (van der Werf, 2009), and are derived mainly from increased accuracy of hard to measure traits including carcass and meat quality, disease resistance, reproduction, and adult wool production. Because many of these traits are typically not measured in industry breeding programs, much of the effort to develop genomic selection for the Australian sheep industry has been in establishing a comprehensively phenotyped reference population to develop genomic predictions. In addition, since many industry ram breeding flocks have been genotyping a proportion of their selection candidates since 2010, there are a large number of animals that can also contribute to genomic predictions for any traits measured in industry flocks. Key industry sires which have been genotyped over time also contribute in a similar fashion.

In this paper, we describe how these resources have been utilised to develop single-step breeding value predictions for most traits in the Sheep Genetics evaluations.

Size of the Reference Populations

The main reference population is the Information Nucleus (IN) flocks were established in 2007 and maintained for 5 years (Van der Werf *et al.*, 2010). These flocks were run at 8 locations representing the major sheep production environments across Australia. Approximately 4,000 predominantly Merino ewes were mated annually to Merino ($n \approx 40$), maternal ($n \approx 20$), and terminal sires ($n \approx 40$) selected on the basis of industry relevance and to represent the diversity of the respective populations. An earlier research flock with phenotypes and genotypes (Oddy *et al.*, 2007) also contributed to the reference population. From 2012, the IN flocks were transitioned to the Sheep Genetics Resource Flock (RF) and continued at two of the original sites, assessing approximately 2,000 animals annually and focusing on lamb growth and carcass traits.

Table 1. Summary of the number of genotyped sheep used in the genetic evaluation of sheep in Australia (July 2017).

Animal Group	Maternal	Merino	Terminal	Total
Industry progeny	2,695	8,248	3,423	14,366
Resource flock progeny	4,530	12,810	11,127	28,467
Industry sires	457	1,695	2,137	4,289
Total	7,682	22,753	16,687	47,122

An extensive recording program was conducted in both IN and RF, including measurement of carcass and eating quality traits on slaughtered progeny, a comprehensive wool measurement program in the IN, growth and ultrasound scanning of muscle and fat, resistance to gastrointestinal nematodes (worm egg count), visually assessed traits for wool and body conformation, and reproduction data collected on female progeny of the Merino and maternal sires. Progeny were genotyped using the Illumina 50K ovine SNP chip initially and subsequently the 15K low density chip (Illumina Inc., San Diego, CA).

In addition to these 2 reference populations, a pool of widely used industry sires with accurate estimated breeding values (EBVs) from the main genetic evaluation analyses were

also genotyped as a resource for validating genomic predictions. At the time of writing, there were 1,617 Merino validation sires, 467 maternal sires, and 2,263 terminal sires.

The significant number of genotyped and phenotyped animals from seedstock breeders now also contribute to the genomic predictions of young animals. The number of progeny in each reference set with both genotypes and phenotypes is shown in Table 1.

Development of Single Step Evaluations

Implementing single-step analyses (SS-GBLUP) enables simultaneous use of pedigree, phenotypes and genotypes (Legarra *et al.*, 2014). The benefits of SS-GBLUP in Australian sheep are that genomic information can be propagated to more animals, the size of reference populations can be increased, assumptions of genomic prediction accuracy are not needed and the potential for double counting inherent in blending methods can be avoided. Sheep Genetics conducts analyses across 3 breed traits (Maternal, Terminal and Merino), such that analyses are across generally similar and often related breeds. To date the use of genomic information has been restricted to the major breeds for which we have sufficient information.

In the SS-GBLUP method, the sub-matrix of the inverse numerator relationship matrix (\mathbf{G}^{-1}) in the mixed model equations for genotyped animals is augmented by \mathbf{G}^{-1} where \mathbf{G} is the genomic relationship matrix (Yang *et al.* 2010) and \mathbf{A} is the numerator relationship matrix for genotyped animals, with the resulting modified inverse relationship matrix denoted as \mathbf{G}^{-1} . Solving the single-step BLUP equations with \mathbf{G}^{-1} is more computationally demanding because of the increased density of equations for genotyped animals. This method has now been implemented in all the multiple-trait analyses for Sheep Genetics. A new solving subroutine was developed using highly optimised matrix libraries, enhanced multi-processing capabilities and optimised pre-conditioners. This new software has made it possible to solve complex analyses with up to 3.1 million animals, 76 traits, 549 genetic groups and with several tens of thousands of genotyped animals in less than 48 hours and being run on a fortnightly basis.

Weighting of pedigree and genomic information

Implementation of single step genomic evaluations requires an assumption as to the weighting of the genomic and pedigree relationships in modelling genetic co-variance. A weighting parameter λ ranging between 0 and 1 can be used in the statistical model, with higher values corresponding to greater weighting of genomic information. McMillan and Swan (2017) investigated appropriate values of λ for a range of carcass traits in terminal sire sheep breeds, using the accuracy of genomic prediction of breeding values as a criterion. The observed genomic prediction accuracy generally increased with λ , although the “optimal” value of λ at the maximum accuracy varied widely, covering almost the entire range of possible values. Accuracy typically approached an asymptote towards the optimal λ , so a wide range of values could be used with minimal loss of prediction accuracy. Estimated breeding values (EBVs) calculated with $\lambda = 0.5$ and $\lambda = 0.95$ were highly correlated, although genotyped animals without phenotypic records showed more variation in EBV among animals when λ was higher. In addition, bias (over-prediction of EBVs) increased with λ but was usually acceptable for mid-range values. As a result of these studies, the routine evaluations for Sheep Genetics currently use a conservative λ value of 0.5.

Estimation of ASBV accuracy

To accompany the implementation of multi-trait SS-GBLUP in the Sheep Genetics evaluation systems, an algorithm to approximate accuracy with genomic information was developed (Li *et al.*, 2017). Data from full terminal sire LAMBPLAN analyses were processed using this new method. The results demonstrate that the approximated accuracy of SS-GBLUP EBVs is highly correlated ($R^2 > 0.96$) with exact accuracies in several small example analyses. As expected, SS-GBLUP EBV accuracies increase more for traits with a larger reference population and for traits with higher heritabilities. Animals with low pedigree-only (ABLUP) EBV accuracies benefit more from genomic information than animals with high ABLUP EBV accuracies. The improvement of SS-GBLUP accuracy were on average 18%, 6% and 0.2% points higher for animals with low (<30%), medium (30 to 50%) and high (>50%) starting ABLUP accuracies, respectively (Li *et al.*, 2017).

Validation of single step breeding values

Accuracies of genomic predictions in Australian sheep populations have been published in a number of studies (e.g. Daetwyler *et al.*, 2012a; 2012b; Moghaddar *et al.*, 2013). Genomic prediction accuracies estimated for most traits have generally been moderate, in the range 0.2 to 0.5. To further test the new SS-GBLUP, ASBVs were validated to assess the change in the accuracy of their prediction of progeny performance by Gurman *et al.* (2018). Cross-validation was used to determine differences in accuracies obtained from ABLUP to SS-GBLUP. The average increase in the accuracy of predicting differences in progeny performance from ABLUP to SS-GBLUP was approximately 0.08 averaged across all traits studied. The largest increases were observed in genotyped animals across all trait groups investigated, including hard to measure traits.

Future Development

While the current analyses converge within acceptable time frames, as the number of genotyped animals grows this will become increasingly challenging for the current SS-GBLUP implementation. It is likely that as the number of animals genotyped new computing strategies and analysis models will be required in the not too distant future (e.g. Fernando *et al.*, 2014). Furthermore as the density of marker information increases alternate models which accommodate differential weights for certain markers, or groups of markers, may become more beneficial.

Single-step breeding values for reproduction traits

ASBVs for female reproduction traits currently do not include any genomic information. However, these traits are classic candidates for genomic selection because they are economically important in breeding objectives, lowly heritable, sex limited, expressed late in life, and not as widely or well recorded (as growth, wool and scan traits) in seedstock flocks in Australia. Developing genomic predictions for reproduction has been difficult precisely because they are hard to measure, and there are limited records in the reference population. Recent studies have demonstrated encouraging improvements in accuracy by using genomic information (Daetwyler *et al.*, 2014; Bolormaa *et al.*, 2017).

Work is currently underway to develop an enhanced reproduction analysis which incorporates genomic information, additional correlated traits, and separating net reproduction rate into its component traits of fertility, litter size and ewe rearing ability at both yearling and adult ages (Bunter *et al.*, 2016).

Genetic grouping and breed composition

Genomic information offers significant potential to define flock, strain and breed based structure in the sheep populations (Brown *et al.*, 2013; Dodds *et al.*, 2013; Swan *et al.*, 2014; Gurman *et al.*, 2017). Using genomic information to define genetic groups is likely to be a useful adaptation of SS-GBLUP models. Agreement between estimated marker and pedigree based genetic group effects is reasonable across a range of traits but not perfect. Particularly in single-step analyses for sheep, it will be necessary to further develop methods to define genetic groups using both pedigree and marker information to ensure that genetic group information from un-genotyped animals can be used alongside genetic group differences derived from genomic data.

Genomic prediction across breeds

The current routine analyses restrict the use of genomic information to the main breeds which have sufficient reference animals and include Merino, Poll Dorset, White Suffolk and Border Leicester breeds. The ability to develop genomic predictions across breeds is highly desirable in all livestock populations, to make best use of available reference populations and to provide predictions for numerically small breeds. At the current 50K SNP density the reference populations are not able to predict across breeds (Daetwyler *et al.*, 2010, Moghaddar *et al.*, 2013), and there is some evidence that including animals from another breed can decrease the accuracy for the target breed. The additional challenge in Australia is the increased prevalence of composite sheep. Prediction of breeding values for crossbred animals has been a focus of research and accuracies are high, as long as the contributing breeds are well represented in the reference. At present, genomic information is only utilised in the analysis for animals that are mostly “pure” for the main sheep breeds with sufficient reference population records. More research is required to explore the use of genomic information from composite animals.

Conclusion

Genomic information has now been directly incorporated into genetic evaluations for Australian sheep on a widespread scale using single-step genomic BLUP. Reference populations, both in research and industry settings continue to provide genomic predictions with moderate accuracy, and genomically enhanced breeding values are now available for most traits. The next challenge is to develop improved single-step analyses that can continue to accommodate larger numbers of genotyped animals into the future.

Acknowledgements

This work was funded by Meat and Livestock Australia, Australian Wool Innovation and the Cooperative Research Centre for Sheep Industry Innovation, made possible through the support of the Australian sheep industry. The authors acknowledge the contributions of the

Sheep CRC Information Nucleus, MLA Resource Flock, the Australia Merino Sire Evaluation Association and industry-funded research flocks, and the contribution of sheep breeders through phenotyping and genotyping.

List of References

- Brown, D.J., A.E. Huisman, A.A. Swan, H.-U. Graser, R.R. Woolaston, A.J. Ball, K.D. Atkins, K.D. & R.G. Banks, 2007. Genetic evaluation for the Australian sheep industry. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 17: 187–194.
- Brown, D.J., A.A. Swan, J.S. Gill & R.B. Banks, 2013. Strategies to objectively group merino flocks in sheep genetics. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 20: 66–69.
- Bolormaa, S., D.J. Brown, A.A. Swan, J.H.J. van der Werf, B.J. Hayes, & H.D. Daetwyler, 2017. Genomic prediction of reproduction traits for merino sheep. *Anim. Genet.* 48 (3): 338–48.
- Bunter, K.L., A.A. Swan, I.W. Purvis & D.J. Brown, 2016. Pregnancy scanning can be used as a source of data for genetic evaluation of reproductive traits of ewes. *Anim. Prod. Sci.* 56 (4): 679–89.
- Daetwyler, H.D., S. Bolormaa, K.E. Kemper, D.J. Brown, A.A. Swan, J.H.J. van der Werf & B.J. Hayes, 2014. Using genomics to improve reproduction traits in sheep. *Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada, 17-22 August. Paper 156.*
- Daetwyler, H.D., J.M. Hickey, J.M. Henshall, S. Dominik, B. Gredler, J.H.J. van der Werf & B.J. Hayes, 2010. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Anim. Prod. Sci.* 50: 1004–1010.
- Daetwyler, H.D., K.E. Kemper, J.H.J. van der Werf & B.J. Hayes, 2012a. Components of the accuracy of genomic prediction in a multi-breed sheep population. *J. Anim. Sci.* 90: 3375–3384.
- Daetwyler, H.D., A.A. Swan, J.H.J. van der Werf & B.J. Hayes, 2012b. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genet. Sel. Evol.* 44: 33.
- Dodds, K.G., S.-A.N. Newman, B. Auvray & J.C. McEwan, 2013. Breeds of New Zealand sheep - as recorded or by genomic prediction. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 20: 274–277.
- Fernando, R.L., J.C.M. Dekkers & D.J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet. Sel. Evol.* 46(1): 50.
- Gurman P.M., A.A. Swan & V. Boerner, 2017. Use of genomic data to determine breed composition of Australian sheep. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 22: 1.
- Gurman, P.M., A.A. Swan, V. Boerner, & D.J. Brown. 2018. Cross validation of single step BLUP applied to Terminal sire sheep in Australia, 11th World Congr. Genet. Appl. Livest. Prod. these proceedings
- Legarra, A., O.F. Christensen, I. Aguilar & I. Misztal, 2014. Single Step, a general approach for genomic selection. *Livest. Sci.* 166: 54–65.
- Li, L., A.A. Swan & B. Tier, 2017. Approximating the accuracy of single step EBVs. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 22: 1.
- McMillan A.J. & A.A. Swan, 2017. Weighting of genomic and pedigree relationships in single step evaluation of carcass traits in Australian sheep. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 22: 1.
- Moghaddar, N., A.A. Swan, & J.H.J. van der Werf, 2013. Genomic prediction of weight and wool traits in a multi-breed sheep population. *Anim. Prod. Sci.* 54(5): 544–549.
- Moghaddar, N., A.A. Swan & J.H.J. van der Werf, 2013. Accuracy of genomic prediction from multi-breed sheep reference population. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 20:220–223.

- Oddy, H., B. Dalrymple, D. Ferguson, R. Forage, N. Cockett, F. Nicholas, J. McEwan, J. Kijas, B.J. Hayes, J.H.J. van der Werf, D. Emery, P. Hynd, T. Longhurst & T. Fischer, 2007. SheepGenomics and the International Sheep Genomics Consortium. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 17:411-417.
- Swan, A.A., D.J. Brown & R.G. Banks, 2017. An update on genetic progress in the Australian sheep industry. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 22: 1.
- Swan, A. A., D.J. Brown, H. D. Daetwyler, B.J. Hayes, M.J. Kelly, N. Moghaddar, & J.H.J. Van der Werf. 2014. Genomic Evaluations in the Australian Sheep Industry. *Proc. 10th World Congr. Genet. Appl. Livest. Prod.*, 334. Vancouver, Canada.
- Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123: 218–223.
- van der Werf, J.H.J., 2009. Potential benefit of genomic selection in sheep. *Proc. Assoc. Adv. Anim. Breed. Genet.* 18: 38-41.
- van der Werf, J.H.J., B.P. Kinghorn & R.G. Banks, 2010. Design and role of an information nucleus in sheep breeding programs. *Anim. Prod. Sci.* 50(12): 998–1003.