Proc. Assoc. Advmt. Anim. Breed. Genet. 26: 138-141

VALIDATION OF GENOMIC PREDICTION FOR METHANE EMISSION IN AUSTRALIAN MERINO SHEEP

M.N. Aldridge[1], S.I. Mortimer[2], D.J. Brown[1], P.T. Fitzgerald[3], E.H. Clayton[2], A. Donaldson[2], B. Paganoni[4], A. Thompson[5], V.H. Oddy[2,3] and J.H. van der Werf[3]

[1] Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia
[2] NSW Department of Primary Industries and Regional Development, Armidale NSW, 2351 Australia
[3] School of Environmental and Rural Science, University of New England, Armidale, NSW, 2351 Australia
[4] WA Department of Primary Industries and Regional Development, Bunbury, WA, 6230 Australia
[5] Centre for Animal Production and Health, Murdoch University, Murdoch, WA, 6150 Australia

SUMMARY

Methane emissions are being recorded in Merino sheep across Australia at a large scale. These records will form a reference population that will be used to estimate genomically predicted breeding values for methane for the sheep industry. This paper aims to determine the predictive ability of the current data to estimate methane rate (ml/min) EBVs using single step GBLUP. Phenotypic data was collected in research and commercial seedstock flocks using portable accumulation chambers. Accuracy of prediction was tested through cross validation, using individual flock data as independent test sets. Accuracy ranged from 0.02 and 0.42 for individual animal EBVs. The large range is likely due to variation in linkage with the reference population, environmental conditions prior to measurement, and some inconsistency of measuring protocol. Further work is needed to improve recording strategies and the size and diversity of the reference population.

INTRODUCTION

Selecting sheep that produce less methane (CH_4) is an achievable goal. High and low sheep selection lines for CH_4 had an average of 10 to 12% difference after 10 years of selection (Rowe *et al*. 2019). The complexity and cost of measuring CH_4 makes it a trait that can benefit from genomic prediction. Current projects are designed to collect enough phenotypes for a reference population to provide the industry with genomically predicted CH_4 breeding values. Additionally, there are several past projects that measured CH_4 on large numbers of animals (Robinson *et al*. 2014; Goopy *et al*. 2016; Paganoni *et al*. 2017; Wahinya et al. 2022; Sepulveda *et al*. 2022). Merino sheep are the most abundant breed in Australia are and consequently the largest CH_4 source in the sheep industry, substantiating a larger representation of Merinos in both recent and historic data. The primary aim of this work was to validate the ability of Merino data recorded in the current Australian projects to predict CH_4 breeding values of industry seedstock flocks. The secondary aim was to determine if the data from earlier projects provides valuable information for improving prediction accuracy.

MATERIALS AND METHODS

All CH_4 rate (ml/min) measurements were recorded using the portable accumulation chamber (PAC) methodology (Goopy *et al*. 2011). The data was divided into data collected during projects

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

from 2021 onwards (recent), and projects from between 2009 and the start of 2021 (historic). The data was split as recording protocols and research flock structures have changed over time. The protocol for recent animals included 5 to 7 experimental runs per day, with 12 animals per run. Before each run the animals to be measured were taken off feed 1 hour before entering the PAC. Ewes were then measured for 40 minutes and lambs for 50 minutes. The historic data had varying numbers of runs per day, number of animals per run, animal ages, and the maximum time in the PAC was up to 1 hour. While CH_4 rate was recorded on multiple breeds, this analysis was limited to only include purebred Australian Merino sheep. The recent data includes 5,628 genotyped animals with CH_4 rate measured, from across 11 flocks which includes both industry and research flocks. The historic data includes 3,616 genotyped animals with CH_4 rate recorded from 7 sites, 3 of which were also used in the recent projects. This historic data provides potential linkage between the two data sources. The flocks were spread across Australia, but were predominately from NSW and WA. The historic data was sourced from both the Information Nucleus Flock and Sheep Genomics Flock, while the recent data was recorded on ewes from the AWI Merino Lifetime Productivity (MLP) project, industry ewes from seedstock breeders, and lambs from the MLA Resource Flock (RF). Due to heterogeneity of the data, CH_4 rate was standardised by flock mean and flock standard deviation.

Univariate animal models via restricted maximum likelihood (REML) were used to estimate all variance components, using WOMBAT (Meyer 2007). The model can be summarised as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{ZQg} + \boldsymbol{Za} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is a vector of trait observations for CH_4 rate (CH_4 ml/min), $\boldsymbol{\beta}$ is a vector of fixed effects including Date x Run x Flock, Birth type x Rear type, Sex, Age, Feed Level (FL) x Flock, **g** is a vector of random genetic group effects based on flock of origin, $\boldsymbol{a} \sim (\mathbf{0}, \mathbf{H}\sigma_a^2)$ is vector of additive genetic effects (within genetic groups), where **H** is a combined **A** (pedigree) and **G** (genomic) relationship matrix, and **e** is a vector of residual effects. The matrices **X**, **Z**, and **Q** are design matrices for the corresponding model effects. There were differences in available feed between flocks, and differences were observed in feeding behaviour of individuals within a flock. Feed level is an estimate of feed intake relative to estimated maintenance during the 24 hours before measuring CH_4. This was obtained from measurements of CH_4, CO_2 and O_2 fluxes in the PAC, calibrated against data where feed intake was known (V.H. Oddy *pers. comm.*). Analyses of the recent data included both the full model and ignored the Feeding Level effect; a combined analysis of recent and historic data did not fit Feeding Level due to lack of data to allow its calculation.

For validation of prediction accuracy, single step genomic best linear unbiased prediction EBVs of all animals in the recent data were estimated (Legarra *et al.* 2014), using a training dataset with all phenotypes but those from one test flock removed, and comparing these EBVs with phenotypes of the test flock animals. This procedure was repeated by using each flock as a test set, and repeated again with one test of 512 random animals (mean number of animals removed per test flock). These phenotypes were regressed on EBVs to estimate prediction bias with an expected slope of 1. The accuracy was estimated as *COR(Predicted phenotype, EBV)*/$\sqrt{h^2}$. To estimate linkage, the mean of the top 10 genomic relationships was calculated between animals in the test set and the reference population (animals in the training set that are both genotyped and phenotyped). The mean of the top 10 genomic relationships between the test set and each flock in the training set was also calculated. A final analysis was conducted to determine the predictive ability of historic data for recent animals. Whereby, all the historic data was treated as the training set and all of the recent data as the test set, and then repeated with the reciprocal training and test set.

RESULTS AND DISCUSSION

The heritability for CH_4 rate using recent data was higher when using the full model (0.19), than when the FL x Flock effect was not fitted (0.17). Fitting FL x Flock mostly decreased the residual variation (0.35 down to 0.32) while additive genetic variance (0.07) and genetic group variances

(<0.01) were unchanged. This change in residual variance shows the importance of measuring CO_2 and O_2 to adjust for different feeding levels before recording CH_4. Alternatively, if more feed intake data becomes available, we recommend further investigation of correcting CH_4 rate measurements for time since last feeding. Using the combined recent and historic data resulted in a lower heritability (0.16), which could be due to the changes in recording practices and the changes to population structure over time.

For some flocks, CH_4 rate EBVs were predicted with high accuracies (Table 1). The validation results were highly variable between test subsets, which can partially be explained by the linkage between flocks (Figure 1). The mean of the top 10 relationships between test flocks and the reference population ranged between 0.35 and 0.40. While the range was small, there was a tendency for higher accuracies to have higher linkage with the reference population. Ewe flocks with higher relationships with MLP or other industry flocks, had higher accuracies than flocks that had lower relationships with other ewe flocks and higher relationships with the lamb flocks. Further investigation is required to determine if CH_4 rate measured on ewes and lambs should be treated as the same genetic trait. Future recording strategies should consider recording CH_4 rate on industry lambs and targeting industry flocks that can help improve linkage between industry flocks. While there were 3 flocks that were poorly predicted with low accuracies (0.02 to 0.06), the majority of flocks had reasonable accuracies (0.14 to 0.42) that would improve genetic selection. As more data becomes available and as the measuring protocol is further refined to reduce environmental errors, it is expected that accuracies will further improve.

The bias was different to the expectation of 1 for most flocks (Table 1). The variation in bias across the different test sets demonstrates that the EBVs were either over- or under-dispersed relative to the variation in progeny performance observed in each flock. These results could potentially be explained by differences in genetic diversity for each flock, environmental differences (especially diet) and measurement error. This needs further investigation, but the analysis would be limited by the number of animals measured per flock in the available data.

The historic data as a training set was a better predictor of recent animals (accuracy of 0.12), than recent data as a training set was for historic animals (accuracy of 0.04). This is beneficial as recent animals are the selection candidates. This analysis should be repeated for other breeds or with a multi-breed evaluation, in which case the historic data could be useful. While the recent data in this analysis only used purebred Merinos, there are over 3,000 animals recorded that represent other breeds and Merino crosses, additional recording of animals that are not Merino are planned for 2025.

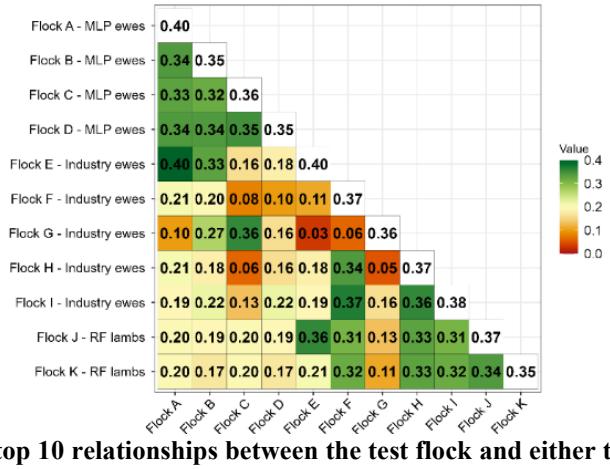

Figure 1. Mean of the top 10 relationships between the test flock and either the full reference population (diagonal) or other individual flocks in the reference population (below diagonal)

Table 1. Validation results from regression of pre-corrected CH_4 rate (ml/min) on EBVs

Test subset	Source of animals in flock	Sires	Animals	Bias	Acc.
Random	Reference pop.	246	512	0.91 ± 0.38	0.41 ± 0.10
Flock A	MLP ewes	29	824	0.60 ± 0.23	0.38 ± 0.08
Flock B	MLP ewes	30	502	0.98 ± 0.29	0.22 ± 0.10
Flock C	MLP ewes	28	575	0.23 ± 0.51	0.18 ± 0.10
Flock D	MLP ewes	25	501	0.30 ± 0.28	0.02 ± 010
Flock E	Industry ewes	46	478	1.43 ± 0.44	0.42 ± 0.10
Flock F	Industry ewes	35	382	1.75 ± 0.60	0.33 ± 0.12
Flock G	Industry ewes	31	484	0.77 ± 0.28	0.22 ± 0.10
Flock H	Industry ewes	45	442	0.59 ± 0.47	0.19 ± 0.11
Flock I	Industry ewes	41	494	0.53 ± 0.47	0.06 ± 0.10
Flock J	RF lambs	109	479	0.19 ± 0.28	0.14 ± 0.11
Flock K	RF lambs	74	467	-0.05 ± 0.41	0.02 ± 0.11

CONCLUSION

Methane traits are expensive to collect, biologically complex, and environmental factors affect the measurements. The validation in this paper indicates a reasonable prediction can be made and used for genetic improvement of Australian Merinos. This prediction depends on linkage to the reference population, which has implications for future recording strategies. The historical data has some value for improving prediction accuracy of recent animals. Further investigation is needed for using both ewe and lamb measurements and for a multi-breed analysis.

ACKNOWLEDGEMENTS

Supported in part by Meat and Livestock Australia (MLA), P.PSH.2011 and L.GEN.2204. Ethics approval; University of New England's Animal Ethics Committee (ARA21-086). With support for the historic data provided by WA-DPIRD, The Australian Federal Government, Murdoch University, MLA, and Australian Wool Innovation. We also acknowledge all the people involved with the data collection.

REFERENCES

Goopy J.P., Woodgate R., Donaldson A., Robinson D.L. and Hegarty R.S. (2011) *Anim. Feed Sci. Technol.* **166**: 219.

Goopy J.P., Robinson D.L., Woodgate R.T., Donaldson A.J., Oddy V.H., Vercoe P.E. and Hegarty R.S. (2016) *Anim. Prod. Sci.* **56**: 116.

Meyer K. (2007) *J. Zhejiang Univ. Sci. B*. **8**: 815

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

Paganoni B., Rose G., Macleay C., Jones C., Brown D.J., Kearney G., Ferguson M. and Thompson A.N. (2017) *J. Anim. Sci.* **95**: 3839.

Robinson D.L., Goopy J.P., Hegarty R.S., Oddy V.H., Thompson A.N., Toovey A.F., Macleay C.A., Briegal J.R., Woodgate R.T., Donaldson A.J. and Vercoe P.E. (2014) *J. Anim. Sci.* **92**: 4349.

Rowe S., Hickey S., Jonker A., Hess M., Janssen P., Johnson T., Bryson B., Knowler K., Pinares-Patino C. and Bain W. (2019) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **23**: 306.

Sepulveda B.J., Muir S.K., Bolormaa S., Knight M.I., Behrendt R., MacLeod I.M., Pryce J.E. and Daetwyler H.D. (2022) *Front. Genet.* **13**: 883250.

Wahinya P.K., Oddy V.H., Dominik S., Brown D.J., Macleay C.A., Paganoni B., Thompson A.N., Donaldson A.J., Austin K., Cameron M. and van der Werf J.H.J. (2022) *Anim. Prod. Sci.* **62**: 818.