SHARING MULTIBREED COW DATA WITH NEW ZEALAND TO IMPROVE PREDICTION FOR AUSTRALIAN CROSSBREED COWS FOR MILK YIELD TRAITS

M. Haile-Mariam[1], I.M. MacLeod[1], M. Khansefid[1], C. Schrooten[2], E. O'Connor[3], G. de Jong[2], H.D. Daetwyler[1,4] and J.E. Pryce[1,4]

[1]AgriBio Centre for AgriBioscience, Agriculture Victoria, Bundoora, VIC, 3083 Australia
[2]CRV, 6800 AL Arnhem, the Netherlands
[3]CRV Ambreed, Hamilton, 3216 New Zealand
[4]School of Applied Systems Biology, La Trobe University, Bundoora, VIC, 3083 Australia

SUMMARY

Ways to improve accuracy of genomic prediction (GP) for Australian (AU) crossbred cows by using data of about 33,000 cows from New Zealand (NZ), where crossbreeds are the dominant breed group (BG), and AU data were assessed. Accuracy of GP for validation cows was tested using single trait and multi-trait models, with data from different BGs considered as correlated traits. When data of the different BGs were considered as separate traits, the genetic correlations for milk yield (MY) were higher compared to that for fat yield (FY). The lowest correlations for all traits were between pure Holstein (H) and Jersey (J) as expected, and among the milk yield traits the lowest correlations were for FY. The estimated heritability and genetic correlations using the high-density SNP chip were slightly higher than 50K chip. Accuracy of GP using the NZ reference set (RS) was not better than AU reference. For MY, the accuracy of GP for AU crossbreed cows was like that observed for pure breed H cows. However, for FY and protein yield (PY), the accuracy of GP was lower in HJ (F_1) and HHJ (back cross to H) crosses. The joint NZ-AU RS resulted in 1 to 5% increase in accuracy for FY and PY of mainly crossbred cows.

INTRODUCTION

A joint project to improve accuracy of GP by sharing cow data in the pasture-based dairy systems of NZ and AU has been established by Agriculture Research Victoria and CRV (cattle breeding company in The Netherlands). A recent analysis showed that reliability (i.e. squared accuracy) of GP for milk traits for NZ validation bulls can be increased by 4 to 7% by including about 60,000 AU cows to a RS that included all NZ animals (Haile-Mariam *et al.* 2019). The benefit of adding NZ cows to AU RS is expected to be low for AU pure breed prediction because the number of genotyped NZ cows is relatively small. However, the number of crossbred cows from NZ is more than that from AU and this could be used to improve accuracy GP for AU crossbreed cows and possibly even for purebred Jersey for which the AU RS is small. Several studies have shown that the accuracy of GP from multibreed RS is not better than single-breed RS particularly when the breeds are distantly related (Calus *et al.* 2018). The inclusion of crossbred animals in the RS could improve the accuracy GP for crossbreds which was reported to be lower than those observed for pure breeds (Khansefid *et al.* 2019) and for all animals by improving the links between the pure breeds.

Data from several breeds for GP have been used in joint analyses in several ways including by considering the same trait recorded in different breeds as correlated traits in multi-trait (MT) model (Calus *et al*. 2018; Karoui *et al*. 2012) or by fitting breed as fixed effect in univariate model (Uni). In the MT model, the marker effects could be assumed to be different in different breed groups (BGs) where performance in J and H and their different crosses are treated as different but correlated traits. Using milk yield traits as response variable, the objectives of this study were: 1) to estimate genomic correlation (r_g) between the same trait measured in different BGs; 2) to assess the accuracy of GP for AU crossbred and purebred validation cows using NZ and AU cows as RS.

MATERIALS AND METHODS

Performance data of about 33,000 NZ genotyped cows and their contemporaries were obtained from NZ and included in the May 2018 genetic evaluation of DataGene for AU dairy cattle. In addition to NZ cows, there were close to 60,000 AU cows in the dataset. All NZ cows and most AU cows were genotyped with low density SNP chips (~ 10K SNP). These genotypes were imputed first to Bovine 50K SNP chip and then to High Density (HD) 800K SNP panels. After edit, in total the HD genotype set included 633,374 SNP and the 50K chip included 40,850 SNPs. The HD and 50K genotypes were used to create genomic relationship matrices (GRMs). The GRM that included J, H and crossbreeds was calculated for NZ and AU reference and validation cows (Table 1) separately and jointly following Yang et al. (2010). The number of cows included in the RS (born before 2011) and cows used for validation (born after 2010) is shown in Table 1. The response variable which were DRP for milk yield traits were analysed using MTG2 (Lee and van der Werf 2016). When all data were considered as the same trait, BG was fitted as fixed effect and in the multi-trait model data of each BG was considered as separate trait. In addition to the 5-trait (BGs) in NZ and 4-trait model in AU, the data from each country were analysed assuming a 3-trait model by combining the back crosses (i.e. HHJ or JJH) into their respective pure BG.

Adjusted accuracy was calculated as correlation between direct genomic breeding values (DGVs) and DRP, divided by the accuracy of the DRP of the validation cows. To ensure that the accuracies were less affected by high relationship among AU reference and validation cows, a cow was included in the validation set if its genomic relationship to the average of the top 10 cows in the RS (Clark *et al.* 2012) was below 0.25. As a result of this, no J cows were used for validation.

Table 1. Number of NZ and AU cows in reference set by breed group (BG: Holstein [H], back cross to H [HHJ], F_1 [HJ], back cross to J [JJH] and Jersey [J]) which their records were considered as different traits (5, 4 or 3 traits), and the number of AU validation cows

Breed group	NZ reference set			AU reference set			AU validation set
	Number	5-Trait	3-Trait	Number	4-Trait	3-Trait	
H	8624	Trait 1	Trait 1	21633	Trait 1	Trait 1	4944
HHJ	10125	Trait 2	Trait 1	1401	Trait 2	Trait 1	965
HJ	8675	Trait 3	Trait 2	1308	Trait 3	Trait 2	344
JJH	1481	Trait 4	Trait 3	-	-	-	-
J	3915	Trait 5	Trait 3	5905	Trait 4	Trait 3	-

RESULTS AND DISCUSSION

Tables 2 and 3 show the proportion of variance explained by the GRM (genomic h^2) in NZ and AU cows for MY and FY when the HD SNP chip was used. The genomic h^2 was the lowest for PY when using NZ data where they varied from 0.14 to 0.18 and 0.15 to 0.19 using the 50K and HD SNP chip, respectively. In the AU data, genomic h^2 for PY were only slightly lower than or similar to that for FY. In all cases the HD SNP chip explained about 2 to 5% more variance than 50K SNP chip (results not shown). Genetic correlations (r_g) among the BGs, when each BG was considered as traits, were lower for FY than for MY (Table 2 and 3). The pattern of r_g for PY was more similar to MY than to FY in NZ data but similar to FY in AU data. Differences in r_g between SNP chips were small, but in general the HD SNP chip showed higher correlations among the BGs than the 50K SNP chip (Table 2 and 3). As expected, r_g had higher standard errors (up to 0.10) than genomic h^2 (up to 0.04). Although the genomic h^2 were higher when AU cow data were used, the standard errors of the genomic h^2 and the r_g, particularly those involving crossbred BG, were higher in AU than NZ cows. Overall the use of 50K SNP chip "correctly" estimated the r_g to be the lowest between J and

H, whereas the HD SNP chip estimated the lowest correlation to be between H and the HJJ (Table 2), though the differences were not significant given the standard errors which were up to 0.15 in AU data. The observation that the HD SNP chip explained more variance than the 50K SNP chip agrees with van den Berg *et al.* (2016), where they found that adding selected sequence variants increased h^2 compared to the 50K SNP chip. Lower r_g between breeds when BGs are considered as traits for FY compared to MY in this study also agrees with other studies (Calus *et al.* 2018; van den Berg *et al.* 2016). Overall that our r_g estimates even between the two pure breeds (H and J) are higher than most literature estimates (van den Berg *et al.* 2016) may be due to some level of crossbreeding between J and H in NZ and AU (de Roos *et al.* 2008; Pryce *et al.* 2011) several generations back or due to similarity in production environment (i.e. pasture-based). The 3-trait model based on NZ and AU data sometimes showed the lowest r_g to be between HJ and J rather than between J and H which was unexpected. This may be due the small sample size and possibly some errors in the BG classification.

Table 2. Genomic h^2 in NZ reference cows of high-density SNP chip (HD) for milk and fat yield on the diagonal (in bold) and genetic correlations between breed groups for milk and fat yields using HD (above diagonal) and 50K SNP chip (below diagonal) in 5-trait model

Breed group	Milk					Fat				
	H	HHJ	HJ	JJH	J	H	HHJ	HJ	HJJ	J
H	**0.30**	0.97	0.85	0.73	0.77	**0.26**	0.97	0.76	0.42	0.49
HHJ	0.96	**0.30**	0.93	0.85	0.85	0.98	**0.23**	0.86	0.65	0.57
HJ	0.83	0.91	**0.31**	0.95	0.87	0.78	0.86	**0.22**	0.85	0.80
HJJ	0.77	0.84	0.92	**0.36**	0.86	0.41	0.63	0.84	**0.24**	0.90
J	0.72	0.82	0.86	0.89	**0.4**	0.47	0.55	0.78	0.88	**0.27**

Table 3. Genomic h^2 in AU reference cows of high-density SNP chip (HD) for milk and fat yield on the diagonal (in bold) and genetic correlations between breed groups for milk and fat yields using HD (above diagonal) and 50K SNP chip (below diagonal) in 4-trait model

Breed group	Milk				Fat			
	H	HHJ	HJ	J	H	HHJ	HJ	J
H	**0.34**	0.96	0.81	0.88	**0.23**	0.96	0.66	0.57
HHJ	0.94	**0.37**	0.88	0.97	0.92	**0.24**	0.65	0.57
HJ	0.78	0.78	**0.40**	0.78	0.63	0.56	**0.35**	0.55
J	0.75	0.91	0.72	**0.43**	0.43	0.58	0.50	**0.26**

The accuracy of GP for HJ and HHJ was higher than H for MY when NZ cows were used as RS (Table 4) because the crossbred cows dominate the set (Table 1). When using AU RS only, accuracy of GP was lower for crosses compared to H for FY (Table 2 and 3) and PY (results not shown) where r_g between the BGs were also lower. The use of NZ cows as a RS is expected to have less contribution for GP of PY because the r_g between performance in NZ and AU is lower (0.60 in H and 0.70 in J, Haile-Mariam *et al.* 2019) compared to both MY and FY and this will likely reduce the benefit of adding NZ RS to improve GP. However, the use of AU+NZ RS increased adjusted accuracy by 1 to 5% (Table 4). Table 4 also shows that considering performance of cows of different BGs in MT or Uni model has little benefit on the accuracy.

Table 4. Adjusted accuracy as correlation between direct genomic breeding values (DGVs) and DRP, divided by the accuracy of the DRP for validation Australian (AU) cows using New Zealand or AU cows in the reference, assuming data of cows from different breed groups to be the same trait (Uni.) or different (multi-traits) models from HD GBLUP

Trait	Breed group	New Zealand			Australia			AU+NZ
		Uni.	3-Trait	5-Trait	Uni.	3-Trait	4-Trait	Uni.
Milk	HJ	0.46	0.45	0.46	0.61	0.61	0.60	0.61
	HHJ	0.43	0.44	0.42	0.59	0.59	0.58	0.60
	H	0.33	0.33	0.35	0.58	0.58	0.57	0.59
Fat	HJ	0.26	0.24	0.22	0.43	0.42	0.42	0.47
	HHJ	0.28	0.26	0.23	0.46	0.46	0.45	0.48
	H	0.29	0.28	0.28	0.55	0.55	0.55	0.56
Protein	HJ	0.36	0.36	0.34	0.40	0.42	0.40	0.44
	HHJ	0.23	0.24	0.24	0.36	0.36	0.36	0.37
	H	0.34	0.33	0.34	0.54	0.54	0.54	0.55

CONCLUSIONS

Although the NZ reference did not provide better GP accuracy for AU crossbreed cows than AU RS, the joint use of AU and NZ RS increased GP for FY in HJ and HHJ cows and for PY in HJ only. In the case of MY accuracy of GP in crosses and H was similar, so adding NZ cows was not beneficial.

ACKNOWLEDGEMENTS

We thank DataGene and DairyNZ for data and Drs Konstantinov and Nieuwhof of DataGene for the DRPs. This study is funded by DairyBio (Agriculture Victoria and Dairy Australia) and CRV.

REFERENCES

Calus M.P.L., Goddard M.E., Wientjes Y.C.J., Bowman P.J. and B.J. Hayes (2018) *J. Dairy Sci.* **101**:4279.

Clark S.A., Hickey J.M., Daetwyler H.D. and van der Werf J.H. (2012) *Genet. Sel. Evol.* **44**:4.

de Roos, A.P.W., Hayes, B.J., Spelman, R.S. and Goddard M.E. (2008) *Genetics* **179**:1503.

Haile-Mariam M., MacLeod I.M., Schrooten C., de Jong G, O'Connor E., Daetwyler H. D. and Pryce J.E. (2019) Submitted to *J. Dairy Sci.*

Khansefid M., MacLeod I.M., Haile-Mariam M., Schrooten C., de Jong G, O'Connor E., Daetwyler H. D., Pryce J.E. and Goddard M.E. (2019) *Proc. Assoc. Advmt. Anim. Breed. Genet.***23**:(in press).

Karoui S., Carabano M.J., Diaz C. and Legarra A. (2012) *Genet. Sel. Evol*. **44:**39.

Lee, S. H, and van der Werf J. H. (2016) *Bioinformatics*. **1;32**(9):1420.

Pryce J. E., B. Gredler, S. Bolormaa, P.J. Bowman, C. Egger-Danner, C. Fuerst, R. Emmerling, J. Sölkner, M.E. Goddard and B.J. Hayes. (2011) *J. Dairy Sci.* **94**:2625.

Van den Berg I., D. Boichard and M. S. Lund. (2016) *Genet Sel. Evol.* **48**:83.

Yang J, Benyamin B., McEvoy B., Gordon S., Henders A., Nyholt D., Madden P., Heath A., Martin N., Montgomery G., Goddard M.E and Visscher P.M. (2010) *Nature Genet.* **42**:565.