Proc. Assoc. Advmt. Anim. Breed. Genet. 23:123-126

BREEDING VALUES OF THE 1000-BULL-GENOME CATTLE ESTIMATED BY DAIRY PLEIOTROPIC VARIANTS

R. Xiang[1,2] and M.E. Goddard[1,2]

[1]Faculty of Veterinary & Agricultural Science, The University of Melbourne, Parkville 3052, Victoria, Australia
[2]Agriculture Victoria, AgriBio, Centre for AgriBiosciences, Bundoora, Victoria 3083, Australia

SUMMARY

The 1000-Bull-Genome (1KBull) project contains whole genome sequence data of thousands of cattle with different breeds from various countries. While most 1KBull cattle do not have phenotypic data, different breeds display distinct phenotype due to artificial and/or natural selections. For example, the milk production of Holstein cattle is expected to be higher than that of Angus cattle. Such expected phenotypic differences between breeds may be useful for validating the informativeness of a set of prioritised variants. Via meta-analysis of GWAS with 17.6 million imputed sequence variants with over 44,000 Australian dairy cattle, we prioritised a set of 92.5K pleiotropic variants associated with multiple traits including milk production, reproduction, management and linear assessment. With these pleiotropic variants, the genomic best linear unbiased prediction (gBLUP) was used to estimate dairy-trait breeding values (gEBV) for 2,334 1KBull cattle (Run 6). Based on principal components analysis, the dairy-trait gEBVs separated the dairy from beef breeds as well as the separation using whole genome sequence data. For individual trait gEBVs in the 1KBull cattle, while milk, protein and fat yield, somatic cell count, stature and angularity were significantly higher in dairy than in beef cattle, the milk protein and fat percentages, muzzle width and teat length were significant lower in the dairy than in the beef cattle. Compared to 1KBull Jersey cattle gEBVs, Holstein cattle had significantly higher milk, protein and fat yield and stature, but significantly lower fat and protein percentages and somatic cell count. Our study provides valuable insights into the genomic prediction of breed differences using within-breed trained equations. Our work also provides alternative validation strategies for prioritised markers.

INTRODUCTION

The 1000-Bull-Genome (1KBull) project collects whole genome sequence data worldwide via donations from consortium members. Since 2012 (Daetwyler *et al*. 2014), the dataset has grown to over 2,000 cattle from more than 100 breeds of *Bos taurus* and *Bos indicus*. Up to 44 million sequence variants have been identified in the 1KBull cattle and these variants are used as the basis for sequence variant imputation in large cattle populations. Large cattle populations with sequence variants have facilitated genome-wide association studies (GWAS) (Bouwman *et al*. 2018) and genomic prediction (VanRaden *et al*. 2017) of complex traits. Here we examine a new use for the 1KBull database; the prediction of trait differences between breeds.

Genomic prediction is usually used to predict differences in breeding value within a breed and it is unknown if it would correctly predict differences between breeds. One of the aims of this paper is to test the ability of within breed genomic prediction to predict differences between breeds. We develop prediction equations within breeds of dairy cattle and combine them with the genotypes of bulls in the 100KBull database to predict the differences between breeds. These predicted breed differences are compared to expectations such as higher milk yield in dairy breeds than in beef breeds.

MATERIALS AND METHODS

The 1KBull data used in this study was part of the Run 6 (http://www.1000bullgenomes.com/). In total the whole genome sequence data of 2,334 *Bos taurus* cattle were used. Dairy and beef cattle breeds and their sample sizes were defined as in Table 1. The defined dairy and beef cattle breeds were used for gEBV comparisons described later on.

Table 1. Sample size of defined dairy and beef cattle breeds

Dairy cattle		Beef cattle	
Holstein	567	Angus	266
Brown Swiss	148	Simmental	225
Jersey	66	Charolais	128
Montbeliarde	54	Limousin	82
Normandy	44	Hereford	75
Finnish Ayrshire	25	Guelph composite	30
Norwegian Red	24	Beef Booster	29
Guernsey	20	Blonde dAquitaine	26
Swedish Red	16	Belgian Blue	16
		Angus Red	6
		Maine Anjou	5
		BraunviehBeef	4

A set of pleiotropic sequence variants (92.5K) associated with 34 dairy traits were identified using Australian dairy bull (N>11,000) and cow populations (N>33,000) and 17.7 million imputed sequence variants with accuracy R2 > 0.4. The detail of the data and the GWAS model used can be found in (Xiang et al. 2019). Briefly, the traits were decorrelated by Cholesky transformation (Xiang et al. 2017). GWAS fitting breed as the fixed effects were conducted for each one of the 34 traits separately in bulls and cows. For the GWAS results of each trait from two sexes, a weighted t value was calculated to combine the variant effects with $t_w = \frac{\frac{B_{bull}}{se_{bull}{}^2}+\frac{B_{cow}}{se_{cow}{}^2}}{\sqrt{\frac{1}{\frac{1}{se_{bull}{}^2}+\frac{1}{se_{cow}{}^2}}}}$ (Xiang *et al*. 2018) where B_{bull} and se_{bull} were the beta and standard error (se) of the bull GWAS and B_{cow} and se_{cow} were the beta and se of the cow GWAS. The weighted t value across traits and variants were analysed by the multi-trait meta-analysis method (Bolormaa *et al*. 2014). Variants with the meta-analysis P-value < 1e-6 and MAF > 0.001 were retained as significant pleiotropic variants.

The genomic best linear unbiased prediction (gBLUP) implemented in MTG2 (Lee and Van der Werf 2016) was used to train prediction equations in the dairy dataset. A genomic relationship matrix (GRM) was calculated from the prioritized pleiotropic variants. Original traits (deregressed proofs) were used to perform gBLUP in Australian bulls and cows. The gBLUP model used was $y = mean + breed_i + a + error$, where y = vector of phenotypes for bulls or cows, $breed_i$ = three breeds for bulls, Holstein, Jersey and Australian Red and four breeds for cows (Holstein, Jersey, Australian Red and MIX), a = polygenic random effects ~N(0, Gσg2) where G = GRM. This estimated the total genetic value of Australian bulls and cows and was followed by the back-solving for the variant solution in the Australian data. Then, the variant solutions were combined with the sequence genotypes to calculate dairy-trait gEBV of the 1KBull cattle.

RESULTS AND DISCUSSION

A principle component analysis (PCA) was carried out on the sequence genotypes of the 1KBull database and dairy-trait gEBVs (Figure 1). Overall, the first PC separated Holstein from other breeds and the 2nd PC separated Angus from other breeds. This may reflect that these two breeds were the most common in the database. The 1st PC of gEBVs (X-axis of the right panel of Figure 1) associated with milk production traits separated some dairy cattle breeds but did not separate beef cattle breeds. This also suggested that the 37 gEBV of dairy traits can be used to distinguish the phenotypic difference between dairy and beef cattle.

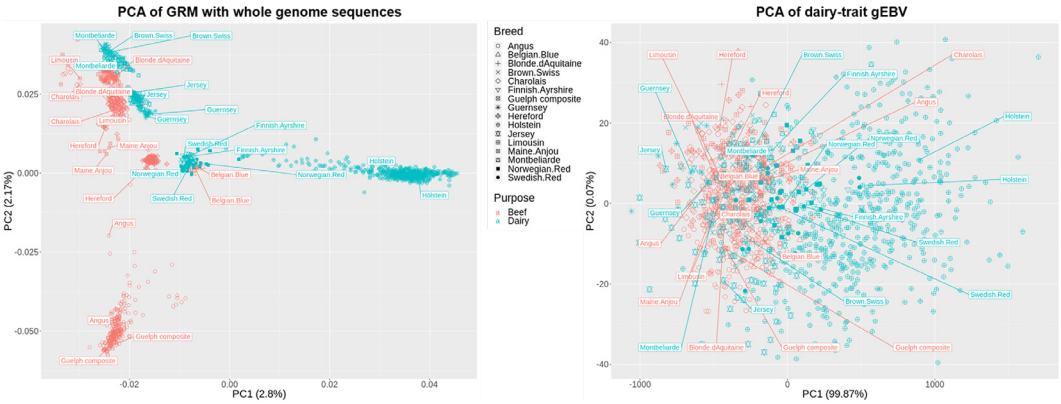

Figure 1. Principal components analysis results of the genomic relationship matrix and the dairy trait gEBVs of the 1000-bull-genome cattle

Individual dairy-trait gEBVs were compared between dairy and beef cattle breeds and were also compared between Holstein and Jersey breeds in the 1KBull individuals (Figure 1 and Table 2).

Table 2. gEBV difference. ns: not significant

gEBVs	Trait full name	Dairy VS Beef	Holstein VS Jersey
Prot	Protein yield	+	+
Fat	Fat yield	+	+
Milk	Milk yield	+	+
FatP	Fat percentage	-	-
ProtP	Protein percentage	-	-
SCC	Somatic cell count	+	-
Temp	Temperament	-	-
MSpeed	Milking speed	+	+
Stat	Stature	+	+
Like	Likeability	-	-(ns)
Angul	Angularity	+	+
MuzW	Muzzle width	-	+(ns)
TeatL	Teat length	-	+(ns)
UdTex	Udder texture	+	+
UdDep	Udder depth	+	+(ns)
RumpL	Rump length	+	+
OType	Overall type	+	+
Mamm	Mammary systems	+	+

Most dairy trait gEBVs were higher in the 1KBull dairy cattle than those in the 1KBull beef cattle. Thus, the within breed genomic predictions do predict qualitative differences between breeds. This result also supports the informativeness of the retained pleiotropic variants. The lower fat (FatP) and protein percentages (ProtP) in the diary breeds than in the beef breeds was due to that their higher milk yield. The somatic cell count (SCC) score and milk speed (MSpeed) was higher in the dairy cattle than in the beef cattle. The dairy cattle are predicted to have better overall type (OType) and mammary system (Mamm), to be more Angular and have shorter teat length (TeatL). These differences appeared to be consistent with the common expectations.

In the gEBV comparisons between Holstein Jersey breeds, Holstein cattle had higher milk productivities, but lower somatic cell count score, fat and protein percentages than Jersey cattle. Holstein cattle had better assessment of the overall type and the mammary system. No significant differences were found for likability, muzzle width (MuzW), teat length and udder depth (UdDep) between the two breeds. These observations appeared to be consistent with the common knowledge about Holstein and Jersey cattle.

CONCLUSIONS

Overall, our results show that it is possible to predict qualitative differences between breeds using genomic prediction based on a set of sequence variants chosen because they are associated with dairy traits. This study also provides alternative insights into efficient use of available data to conduct validation analysis. Our analysis included ~900 beef cattle from the Run6 of the 1KBull project. It is recommended to extend such genomic prediction analysis in a large beef cattle population where the allele frequency of the prioritised dairy pleiotropic variants can be properly examined and accounted for.

REFERENCES

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., Barendse W., Kemper K., *et al*. (2014) PLOS *Genetics* **10**: e1004198.

Bouwman A.C., Daetwyler H.D., Chamberlain A.J., Ponce C.H., Sargolzaei M., Schenkel F.S., *et al*. (2018) *Nature Genetics* **50**: 362.

Daetwyler H.D., Capitan A., Pausch H., Stothard P., Van Binsbergen R., Brøndum R.F., *et al*. (2014) *Nature Genetics* **46**: 858.

Lee S.H. and Van der Werf J.H. (2016) *Bioinformatics* **32**: 1420.

VanRaden P.M., Tooker M.E., O'Connell J.R., Cole J.B. and Bickhart D.M. (2017) *Genet. Sel. Evol.* **49**: 32.

Xiang R., Berg I.v.d., MacLeod I.M., Hayes B.J., Prowse-Wilkins C.P., Wang M., *et al*. (2019) *Proc. Nat. Academy of Sciences* 201904159.

Xiang R., Hayes B.J., Vander Jagt C.J., MacLeod I.M., Khansefid M., Bowman P.J., *et al*. (2018) BMC *Genomics* **19**: 521.

Xiang R., MacLeod I.M., Bolormaa S. and Goddard M.E. (2017) *Scientific Reports* **7**: 9248.