Contributed paper

RANKING BRAHMAN BULLS FOR FEMALE REPRODUCTIVE PERFORMANCE IN NORTHERN AUSTRALIAN COMMERCIAL ENVIRONMENTS USING DNA POOLING

Y. Li[1], L. Porto-Neto[1], R. McCulloch[1], S. McWilliam[1], P. Alexandre[1], J. McDonald[2], A. Reverter[1] and S. Lehnert[1]

[1] CSIRO Agriculture and Food, St Lucia, QLD, 4067 Australia
[2] MDH Pty Ltd, Cloncurry QLD 4824, Australia

SUMMARY

Female fertility is one of the important reproductive traits that directly impact the profitability of commercial beef breeding herds. DNA pooling of cows with reproductive records can provide a cost-effective way for assessing and predicting the contribution of individual bulls to the fertility of their female offspring. However, panels of different SNP density exist and their impact on genomic prediction is unknown when DNA pooling is applied. In this study, using the genotype and phenotype (pregnancy test and lactation status) from two Brahman cattle populations in north Queensland, one containing 715 samples genotyped with 54,791 SNPs, the other consisting of 290 samples genotyped with 74,584 SNPs, we investigated genetic relationships between the two populations as well as rankings of individual bulls based on genomic prediction for pregnancy test outcome of their progeny. Our results show different outcomes obtained from using different density SNP panels in separating cow pooling samples, and estimating genomic breeding values for pregnancy test outcome of individual bull's progeny. The research highlights that extreme caution needs to be taken for choosing SNP panels of different densities to rank and select bulls for commercial beef production based on DNA pooling technology.

INTRODUCTION

Genomic prediction of breeding values based on a genomic relationship matrix has revolutionized the ability to identify genetically superior livestock for improving traits that are difficult to measure (van der Werf 2009). However, in commercial herds, it is impractical to individually genotype all animals. DNA pooling of cows with reproductive records can provide a cost-effective way for assessing and predicting the contribution of individual bulls to the fertility of their female offspring (Reverter *et al*. 2016). A question that remains to be answered is what density SNP panel should be used to genotype DNA pooled cows to rank bulls to achieve accurate prediction of reproductive performance of their progeny? In this study, using two Brahman cattle populations in north Queensland, we aimed to investigate the impact of SNP panels of different density on the ranking of bulls.

MATERIALS AND METHODS

Animals. Datasets from two Brahman cattle populations in north Queensland were used for the study. One (SmartF) consists of 290 samples from 2012-2014 herds (177 individual bulls and 113 pools representing 2,648 cows) genotyped with 74,584 SNPs (770K BovineHD BeadChip platform). The other (MDH2020) contains 715 samples from the 2020 herd (482 individual bulls and 233 pools representing 2,452 cows) genotyped with 54,791 SNPs (Neogen Australasia GGP TropBeef 50K chip). DNA pools were formed based on the pregnancy test (i.e. not pregnant or pregnant) and lactation status (dry or wet) of cows at 2nd joining. Details of the phenotype of pregnancy test outcome (PTO) and pooling techniques can be found in Reverter et al. (2016). In brief, animals were separated into 6 categories, that is, dry and empty (not pregnant, scored as 1), dry and early pregnant (scored 2), dry and mid pregnant (scored 3), dry and late pregnant (scored 4), wet and empty (not pregnant, scored as 5), and wet and pregnant (scored 6). DNA samples of

animals with identical phenotypic scores were pooled together. The individual pool size ranged from 4-45 animals for SmartF (Reverter *et al*. 2016) and from 5-12 animals for MDH2020, depending on the number of animals available in each category. Details of the two datasets are presented in Table 1.

Table 1. Composition of two genotyped populations

Population	Sex	Year	DNA samples	Total
SmartF	Cows	2012	41 (pools)	
(74,584 SNPs)		2013	31 (pools)	
		2014	41 (pools)	113
	Bulls	2013	27	
		2014	150	177
MDH2020	Cows	2020	233 (pools)	233
(54,791 SNPs)	Bulls	2020	482	482

Imputation of genotypic data. Between the two populations, there were 19,089 SNP in common. The imputation from low to high-density SNP genotypes was conducted to both SmartF and MDH2020, using 730,000 SNPs from 5,040 Beef CRC Brahman cattle as the reference. PLINK (Change *et al*. 2015) and Eagle v2.4.1 (Loh *et al*. 2016) were applied for phasing and imputation, respectively. After quality checks with the threshold of R-square value >0.8 and removing SNPs on the sex chromosome, this resulted in 615,310 SNPs.

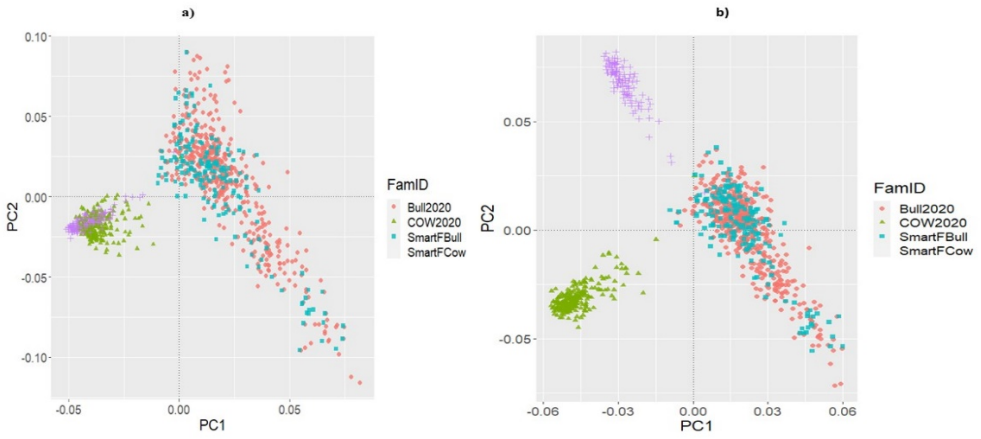
Principal Component Analysis (PCA). To visualize genetic relationships between two populations, we conducted a PCA using genotypes from either the low density (19,089 common SNP) or imputed high-density panel (615,310 SNP, HD).

Genomic prediction. Genomic estimated breeding values (GEBVs) of PTO of progeny for individually genotyped bulls were derived within each population. The conventional genomic prediction method was applied to derive GEBVs, that is, a mixed animal model was used by fitting a polygenic random effect with the GRM (genomic relationship matrix). The fixed effects included the size of pool (30 levels) and contemporary group (5 levels) for SmartF, and SNP chip row (3 levels) and column (24 levels) information for different pools in MDH2020, respectively. The GRM was constructed using the method described by Reverter *et al*. (2016). In brief, the B-allele frequencies from the genotypes of the pools of cows ($\leq$0.25, >0.25 and <0.75 or $\geq$0.75, best fitted the three genotypes based on the individual DNA samples and the genotype call algorithm employed by Illumina) were converted into the three possible genotypes (i.e. 0, 1 and 2 for AA, AB, and BB, respectively) and these were merged with the individual genotypes of each bull to generate a single GRM relating bulls with pools of cows. Then the Qxpak5 software program (Pérez-Enciso and Misztal 2011) was used to fit the GRM in a mixed animal model and obtain genomic estimates of variance components and genomic predictions (GEBVs) for PTO of the testing population. For comparison purposes of different density panels within populations, GEBVs were derived using four GRMs, either with 19,089, 54,791 (for MDH2020 only), 74,584 (for SmartF only), or high density (HD) SNP.

RESULTS AND DISCUSSION

Relationships between animals of two populations. The results from the PCA on all 1,005 animals (290 from SmartF and 715 from MDH2020) are shown in Figure 1. When a low-density SNP panel data (19,089, Figure 1a) was used, 346 DNA pooled cow samples from both populations were clustered together with very small variation among them, suggesting high similarity in the number of alleles between pooled samples. For the 659 individually genotyped bulls (red and blue

dots), there was a much wider range of variation than for cows. However, when the high-density SNP panel was applied (HD, Figure 1b), there was a clear separation of cow samples of within and across two populations. But bulls remained mixed up as low-density results show, with a much narrower range of variation. This indicates that the bulls in the two populations had some degree of relatedness among themselves, but not among the cows. Therefore, the separation of pooled cows would not have been detected if the HD was not used.



individually genotyped bulls (Bull2020), 233 were pools of cow DNA samples (COW2020), 177 were individually genotyped SmartF bulls (SmartFBull) and 113 were pools of SmartF cows (SmartFCow). a) 19,089 common SNP; b) High density SNP

Genomic predictions of bull's PTO with different panels of SNP density. Assuming the results from HD are true, Table 2 shows the Pearson's correlations among the PTO GEBVs from three SNP panels (19,089, 54,791 and HD) in the MDH2020 and SmartF respectively. Within MDH202, the correlations between GEBVs of PTO of 482 bulls were 0.74 between 19,089 and HD, and 0.82 between 54,791 and HD. The correlations were much lower (0.39 and 0.45 respectively) if only the top 25% bulls were considered (see Table 2 correlation for top 25%). Similar trends were observed in SmartF when the correlations of GEBVs for 177 bulls were compared (Table 2), despite slightly higher correlations between 19089 and 74584 with HD when the top 25% bulls were selected (0.54-0.59, Table 2). These suggest that if low-density panels were used to genotype pooled DNA cows for estimating the EBVs of PTO of bulls, at least 40-50% of the best bulls would not be selected.

When further investigating the bull GEBVs of PTO estimated using HD, Table 3 illustrates the profiles of the GEBVs of 482 MDH2020 bulls in different quartiles. The average GEBV difference between top and bottom 25% of bulls was 0.292**,** which is much larger than the difference obtained using low-density panels (0.120 from 19,089 or 0.158 from 54,791, results are not shown here). For animals being dry and empty (score 1) to become wet and pregnant (score 6), there could take conservatively up to 21-27 months to achieve. The GEBV difference of 0.292 from HD would translate into earlier conception by 1.31 months for the female progeny of the top 25% sires.

The study presents preliminary results for the comparison of different panels of SNP density in ranking commercial bulls in two populations. The phenotype score (1-6) of the 2nd joining pregnancy test outcome was treated as a continuous trait in which wet and non-pregnant was scored as "5"

instead of "2". Further research is underway to explore the impact of different score systems on ranking differences.

Table 2. Pearson's correlations among GEBVs estimated from using 19,089, 54,791 and HD SNP panels within MDH2020 and SmartF populations, respectively

Population		MDH2020			SmartF		
	SNP	19089	54791	HD	190	74584	HD
All bulls	19089	1	0.90	0.74	1	0.76	0.72
	54791 / 74584		1	0.82		1	0.81
	HD			1			1
Top 25%	19089	1	0.81	0.39	1	0.52	0.54
	54791 / 74584		1	0.45		1	0.59
	HD			1			1

Table 3. Average genomic breeding values (GEBVs) of progeny pregnancy testing outcome (PTO) of the MDH2020 bulls in four quartiles using HD SNP panel

Quartile	# Bulls	Av. GEBV	Min	Max
1 -Top 25%	120	0.136	0.0833	0.323
2	120	0.055	0.0275	0.0831
3	121	-0.004	-0.0341	0.0261
4 – Bot. 25%	121	-0.156	-0.2771	-0.0345
All	482	0.023	-0.277	0.323

CONCLUSION

This research highlights the need for extreme caution to be taken when applying SNP panels of low or medium densities to study genetic relationships, and rank and select top bulls for commercial beef production based on DNA pooling technology.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Meat and Livestock Australia (MLA) for their financial support to the project P.PSH.1211 "Validation of pooled DNA gEBV for Brahman commercial cow fertility". We also like to thank Clint Smith, manager of Iffley Station of MDH Pty Ltd, for his great assistance in collecting samples and phenotypic information for the study.

REFERENCES

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M. and Lee J.J. (2015) *GigaScience*, **4**.

Loh P.-R., Palamara P.F. and Price A.L. (2016). *Nat. Genet.* **48**: 811.

Reverter A., Porto-Neto L.R., Fortes M.R., McCulloch R., Lyons R.E., Moore S., Nicol D., Henshall J., and Lehnert S.A. (2016). *J Anim Sci*. **94**: 4096.

van der Werf J. H. J. (2009). *Proc. Assoc. Advmt. Anim. Breed. Genet.* **18**:38.