Contributed paper

EVALUATION OF HAPLOTYPE DIVERSITY OF AUSTRALIAN BEEF POPULATIONS USING MEDIUM-DENSITY SNP GENOTYPES

M.H. Ferdosi[1], N.K. Connors[1] and M. Khansefid[2]

[1]Animal Genetics & Breeding Unit, University of New England, Armidale, NSW, 2351 Australia.
[2] AgriBio Centre for AgriBioscience, Agriculture Victoria, Bundoora, VIC 3083, Australia

SUMMARY

Haplotypes as combinations of multiple markers have more diversity than single markers in the population. In this research we studied the haplotype diversity in four beef breeds (Brahman, Hereford, Santa Gertrudis and Wagyu) in Australia to identify the frequent and rare haplotypes within and between breeds. We found that most of the haplotypes (>90%) with more than one percent frequency within each breed were observed in the other breeds as well. Further, the low within breed haplotype diversity in Wagyu can indicate lower genetic diversity compared to the other breeds.

INTRODUCTION

Haplotypes could be more informative of genetic diversity than single markers. However, in genomic predictions, defining relationships between individuals using single markers (VanRaden 2008) is more common than by use of haplotype (Hickey *et al.* 2013; Ferdosi *et al.* 2016) for both single and multi-breed genomic evaluations (Khansefid *et al.* 2020). Phasing of the genotypes into haplotypes and partitioning the genome to multiple segments has several benefits. The accuracy of genomic prediction can be increased using haplotypes instead of single markers (Ferdosi *et al.* 2016; Karimi *et al.* 2018). For example, haplotypes have more diversity than single nucleotide polymorphisms (SNPs). Quantitative trait loci (QTL) can be explored better using haplotypes because crossing over between SNPs and QTL can change the linkage disequilibrium (LD) between them across generations. Consequently, the lower relationships between individuals of different breeds could be precisely defined by calculating the proportion of common haplotypes, which is particularly important in multi-breed genomic predictions. In order for haplotypes to be useful in multi-breed genomic prediction, an overlap of haplotypes across breeds is required. Additionally, haplotypes can be used to calculate genomic inbreeding and provide better insight of relationships between individuals of different breeds. This research investigates the overlap of haplotypes across breeds and their use in the calculation of inbreeding and across-breed relationships.

MATERIALS AND METHODS

Genomic data. Genotypes of four beef breeds in Australia were used in this study to assess the haplotype diversities within and across breeds. The individuals with SNP density greater than 30k SNPs were extracted after quality control and before imputation using the BREEDPLAN genomic pipeline (Connors *et al.* 2018; Johnston *et al.* 2018). Included were 12,692 Brahman with 143,829 SNPs, 21,069 Hereford with 51,441 SNPs, 3,563 Santa Gertrudis (SG) with 82,990 SNPs and 59,120 Wagyu with 51,330 SNPs. No SNPs were removed for low minor allele frequencies (MAF) as these SNPs were important for breed distinction.

Merging the genotypes of four breeds. Genotypes of these four breeds were combined with custom C++ programming to yield 96,444 individuals with 227,422 unique SNPs.

[1] AGBU is a joint venture of NSW Department of Primary Industries and University of New England

Imputation. FImpute Version 2.2 with default parameters (Sargolzaei *et al.* 2014) was used to impute missing genotypes using a multi-breed reference but including the pedigree information. The pedigree had 7% and 4% missing sire and dam, respectively. Prior to imputation within each breed SNPs were removed with missing rate greater than 10% resulting in 29,570 SNPs passing this filter and being used in this study.

Haplotype partitioning. The phased genotypes were partitioned to the haplotype segments with a length of 10 SNPs without overlap (Ferdosi *et al.* 2016). The total number of unique haplotypes within each breed and number of common haplotypes between breeds were calculated using three scenarios: all haplotypes (ALH), haplotypes with frequencies greater than 1 per cent within breed (High-Frequency Haplotype - HFH) and haplotypes with frequencies less than 1 per cent within breed (Low-Frequency Haplotype - LFH). Further, the percent of individuals covered by each of these scenarios were reported. The haplotype diversities/frequencies in the four studied breeds were plotted and analysed using R (R Development Core Team 2020).

RESULTS AND DISCUSSION

Table 1. Number and percent of haplotypes within and between breeds and their population coverage

	Common Haplotypes (A)				**Mean of percent of population covered by common haplotypes (B)**			
	Brahman	Hereford	Santa Gertrudis	Wagyu	Brahman	Hereford	Santa Gertrudis	Wagyu
All haplotypes (ALH)								
Brahman	736,996	55%	48%	45%	100%	78%	94%	81%
Hereford	59%	690,560	47%	49%	94%	100%	91%	86%
Santa Gertrudis	74%	67%	479,930	57%	97%	90%	100%	88%
Wagyu	61%	62%	50%	542,637	87%	92%	87%	100%
Haplotypes with frequency greater than 1% (HFH)								
Brahman	54,052	**92%**	**93%**	**95%**	63%	53%	63%	58%
Hereford	**92%**	54,094	**92%**	**96%**	67%	69%	67%	64%
Santa Gertrudis	**94%**	**93%**	53,535	**95%**	53%	51%	53%	52%
Wagyu	**91%**	**91%**	**90%**	56,684	82%	86%	82%	**93%**
Haplotypes with frequency less than 1% (LFH)								
Brahman	682,944	52%	44%	41%	37%	25%	32%	23%
Hereford	56%	636,466	43%	45%	26%	31%	24%	22%
Santa Gertrudis	71%	64%	426,395	52%	44%	39%	47%	36%
Wagyu	58%	59%	45%	485,953	5%	5%	5%	7%

(A) The diagonal elements are the total number of unique haplotypes in each breed. The non-diagonal elements are percentage of common haplotypes between each pair of breeds, where upper (and lower) triangular elements are number of common haplotypes between breed divided by number haplotypes of the breed in that "row" (and column). **(B)** The diagonal elements show the percentage of the genome covered with haplotypes in each breed. The non-diagonal elements are percentage of genome covered with common haplotypes between each pair of breeds, where upper (and lower) triangular elements are the percentage of the genome of the breed in that "row" (and column) covered with common haplotype.

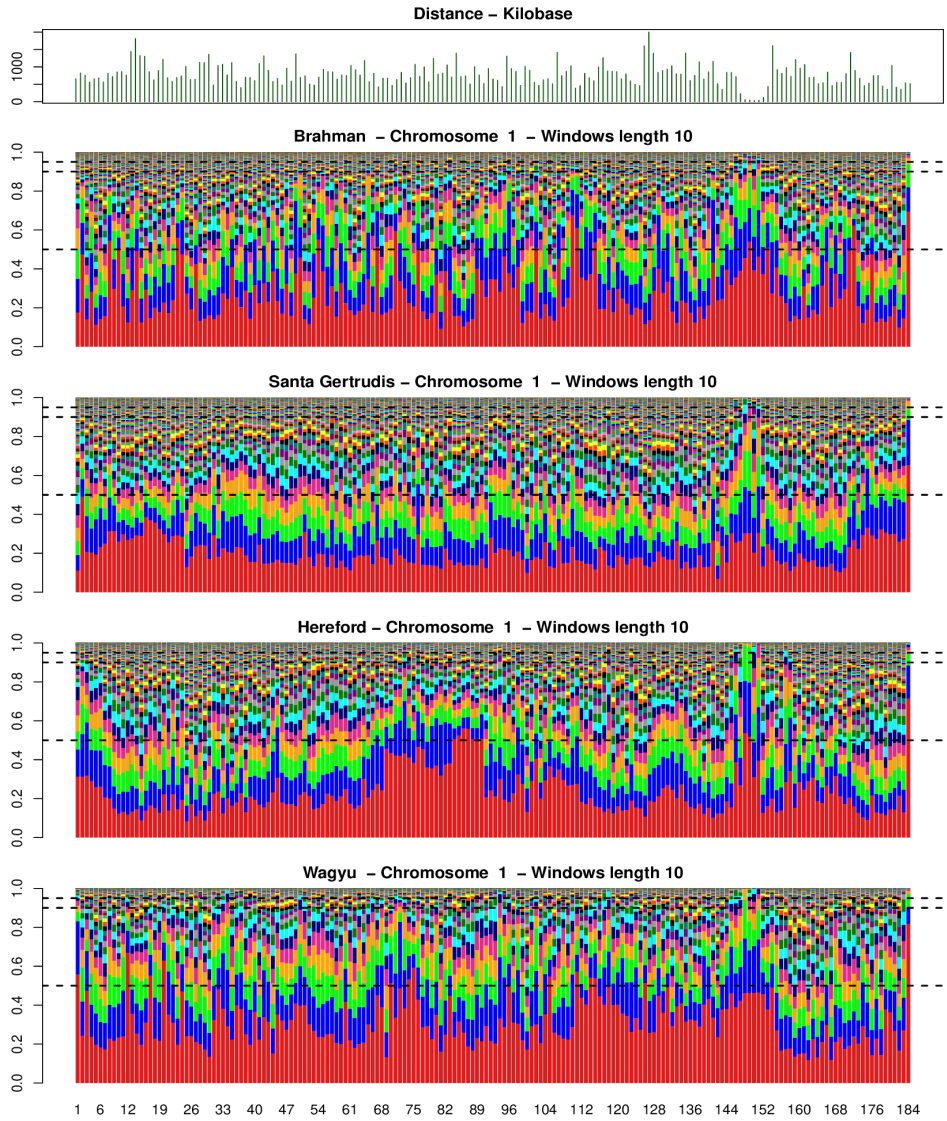

Figure 1 Haplotype diversity across four breeds in chromosome 1. The top plot shows the length of haplotypes (base pair) which were constructed by every 10 adjacent SNP. In the rest of bar plots, each bar represents the sorted haplotype frequencies (vertical line). The horizontal dashed lines mark 50%, 90% and 95% haplotype frequencies. The last haplotype was constructed by less than 10 SNPs, and therefore had fewer haplotypes.

The diagonal elements in Table 1–A shows the total number of unique haplotypes in each of the four breeds with Brahman having the highest number. Santa Gertrudis had the smallest number of unique haplotypes but also had the smallest number of genotyped individuals compared to the other breeds, especially Wagyu. Given the imbalance between number of individuals across breeds could affect our results, the haplotype diversity in Wagyu was especially low which could be a reflection of low effective population size due to limited founders originally imported into Australia. Moreover, according to Table 1-B frequent haplotypes (i.e. HFH) in Wagyu cover 93% of the population while such haplotypes cover 53% of the SG population. The number of common HFH between breeds was very high (Table 1-A-HFH) which can imply the potential in using haplotypes

to improve multi-breed genomic prediction accuracy. Previous studies have already reported the improvement in Restricted Expectation Maximum Likelihood and accuracy of genomic predictions in cross-validation studies using haplotype-based genomic relationship matrices (Ferdosi *et al.*2016). A similar study using a multi-breed population can shed light on the benefits of using haplotypes instead of single markers in multi-breed genomic prediction as our initial haplotype diversity study indicates noticeable overlap between haplotypes in different breeds. The main supporting reason for the usefulness of using haplotypes in multi-breed genomic predictions is the high possibility that many of the QTL and markers are in different LD or even different phase in different breeds. Hence, haplotypes as a combination of multiple markers, could better track QTL especially from distant ancestors compared to single markers. However, it is also important to assess if the haplotypes have the same effects across different breeds.

Figure 1 shows the haplotype diversity in chromosome 1 across the four beef breeds and demonstrates quite different diversity of haplotypes across chromosome 1, which was seen in other chromosomes as well (not shown). As we expected, the marker distance and the length of haplotype significantly affected the haplotype diversities. For example, close to the end of chromosome 1, the lengths of haplotypes were relatively smaller than the rest of haplotypes which could be a potential reason for the lower haplotype diversity in such regions. Possibly partitioning the genome to haplotypes with relatively equal length or recombination rate instead of using similar number of SNPs in haplotypes, could resolve this issue.

CONCLUSIONS

In this study we explored the haplotype diversity within and between breeds. We found that low haplotype diversity within a breed could be an indication of lower genetic diversity in the population such as Australian Wagyu. The haplotype diversity showed relatively high relatedness between different breeds which suggests the potential benefits of using haplotype-based relationships in multi-breed genomic predictions. Further study is required to evaluate the benefits of haplotypes on single marker for genomic prediction.

ACKNOWLEDGEMENTS

MHF, NKC and this study was supported by Meat and Livestock Australia project L.GEN.1704. The authors thank the Hereford, Brahman, Wagyu and Santa Gertrudis societies for providing data for this study.

REFERENCES

Connors N., Cook J., Girard C., Tier B., Gore K., Johnston D., Ferdosi M. (2017) *Proc.Assoc. Advmt Anim. Breed. Genet.* **22**: 317.

Ferdosi M.H., Henshall J., Tier B. (2016) *Genet Sel Evol* **48:** 75.

Hickey J.M., Kinghorn B.P., Tier B., Clark S.A., van der Werf J.H.J., Gorjanc G. (2013) *J Anim Breed Genet* **130:** 259.

Johnston D., Ferdosi M., Connors N., Boerner V., Cook J., Girard C., Swan A., Tier B. (2017) *Proc. Assoc. Advmt. Anim. Breed.Genet.***11**: 269.

Karimi Z., Sargolzaei M., Robinson J.A.B., Schenkel F.S. (2018) *Can J Anim Sci* **98:** 750.

Khansefid M., Goddard M.E., Haile-Mariam M., Konstantinov K.V., Schrooten C., de Jong G., Jewell E.G., O'Connor E., Pryce J.E., Daetwyler H.D., MacLeod I.M. (2020) *Front Genet* **11:** 598580.

R Development Core Team (2020) 'A language and environment for statistical computing. R Foundation for Statistical Computing.' (R Foundation for Statistical Computing:

Sargolzaei M., Chesnais J.P., Schenkel F.S. (2014) *Bmc Genomics* **15:**

VanRaden P.M. (2008) *J. Dairy Sci.* **91:** 4414.