DETECTION OF SIGNATURES OF SELECTION IN AUSTRALIAN BEEF CATTLE

H. Aliloo[1], B.J. Walmsley[2,3], K.A. Donoghue[4] and S.A. Clark[1]

[1]School of Environmental and Rural Science, University of New England, Armidale, NSW 2351 Australia
[2] Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia
[3]NSW Department of Primary Industries, Livestock Industries Centre, Armidale, NSW, 2351 Australia
[4]NSW Department of Primary Industries, Agricultural Research Centre, Trangie, NSW, 2823 Australia

SUMMARY

The 50K genotypes of 2,935 animals from the 5 most common temperate beef breeds in Australia were used to identify genomic footprints of selection based on fixation index (F_{ST}). A principal component analysis on genomic relationships between all individuals showed that Angus, Hereford and Wagyu are the most genetically differentiated breeds. Therefore, 3 pairwise F_{ST} comparisons were implemented between Angus vs. Wagyu, Angus vs. Hereford and Hereford vs. Wagyu. Genome-wide comparison of patterns of the F_{ST} values revealed 14 candidate regions under selection on chromosomes 2:6, 8, 12, 13, 20, and 24. Several of the identified candidate regions in this study have been previously reported for different economically important traits in beef cattle. In addition, our identified candidate regions for signatures of selection harboured genes in several enriched annotation clusters. If validated, the results from this study can be incorporated in genomic selection of the Australian beef cattle population.

INTRODUCTION

Understanding the genetic architecture of productivity is necessary for designing efficient breeding programs. Intensive artificial selection to increase profitability in Australian beef breeds has generated distinctive patterns at specific regions of their genome, referred to as signatures of selection (SoS). The identification of SoS may help to uncover genes and biological mechanisms responsible for breed differences in the Australian beef cattle population.

A simple, yet effective, approach to identify SoS is to compare differences between breeds in allele frequencies of their genome-wide single-nucleotide polymorphisms (SNPs) based on the fixation index (F_{ST}). A high F_{ST} value indicates large differences between the breeds of interests resulted from distinctive selection pressures. Therefore, the comparison of genome-wide patterns of F_{ST} values can help to map genomic regions contributing to the phenotype differences between Australian beef cattle breeds.

The Southern Multibreed (SMB) project has generated genomic data across the 5 most common temperate beef breeds in Australia including Angus, Charolais, Hereford, Shorthorn and Wagyu (Walmsley *et al*. 2021). The aim of this study was to use the genotypes collected in the SMB project to detect SoS in temperate Australian beef breeds using the F_{ST} method.

MATERIALS AND METHODS

Data. The genotypes of 2,938 animals were obtained by Zoetis ZBU medium density 50K (Zoetis, Kalamazoo, MI). The genotype calls with a score of <0.15 were assumed as missing (Edriss *et al.* 2013). Further quality control was undertaken using PLINK 1.9 (Chang *et al.* 2015) to remove

* A joint venture of NSW Department of Primary Industries and University of New England.

SNPs and animals with a call rate lower than 90%, SNPs that were monomorphic across all animals and those located on sex chromosomes (X and Y). Finally, 47,264 SNPs and 2,935 animals including 845 Angus, 493 Charolais, 495 Hereford, 623 Shorthorn and 479 Wagyu cattle were used in this study. The mapping information for all markers was available on the basis of ARS-UCD 1.2 bovine genome assembly.

Data Analysis. To investigate the population structure of different beef breeds, a principal component (PC) analysis based on a genomic relationship matrix (GRM) constructed using GCTA 1.94.1 (Yang *et al.* 2011) was implemented. The first and second PCs were plotted to visualize the distribution and explore the relationships among different beef cattle breeds.

The F_{ST} values were calculated by comparing the allele frequencies of pairwise SNPs between the breeds that showed the highest genetic differentiation based on the first two PCs. PLINK 1.9 (Chang *et al.* 2015) was used to calculate F_{ST} values according to the Weir and Cockerham (1984) method. To reduce the noise in estimates, and to account for linkage disequilibrium between adjacent SNPs, the '*runmed*' R function was used to smooth F_{ST} values across a moving window of 75 markers within each chromosome (Haerdle and Steiger 1995). The SNPs with smoothed F_{ST} values that were greater than 3 times the standard deviation from the mean of all smoothed F_{ST} values (the suggestive threshold) were deemed as being under selection pressure. A candidate region for SoS was defined by first identifying SNPs under selection and then searching within the 500-Kbp interval downstream and upstream (1 Mbp window) of the identified SNP for SNPs that passed the suggestive threshold. The detected region (with a 500-Kbp step size) was extended until there was no SNP with an F_{ST} value greater than the suggestive thresholds within the 500-Kbp interval from the last identified SNP. The boundaries of the candidate region were determined based on the base pair positions of the last-identified SNP in each direction. To visualize the distribution of F_{ST} values across the genome, Manhattan plots were created using the qqman 0.1.4 (Turner 2014) R package. The cattle Quantitative Trait Loci (QTL) database (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index) was used to compare our identified candidate regions to literature. The candidate regions were further investigated for identification of genes residing in them using the biomaRt 2.46.3 (Durinck *et al.* 2009) R package. The identified genes were compared to the whole bovine genome background using functional annotation clustering by DAVID 2021 online bioinformatics resource (Huang *et al.* 2009) to find the biological pathways that are significantly overrepresented.

RESULTS AND DISCUSSION

Figure 1 illustrates that all animals were clearly clustered within their respective breed based on the first two PCs. The PC1 explained around 7.5% of total variation in the GRM and separated Angus and Wagyu breeds from other breeds, while PC2 explained around 5.5% of variation and showed Hereford is genetically more different to Angus and Wagyu than to the other breeds. The Shorthorn and Charolais seemed to be genetically closer to each other based on both PC1 and PC2.

Based on the results from the PC analysis, F_{ST} values were calculated between Angus vs. Wagyu (AW), Angus vs. Hereford (AH) and Hereford vs. Wagyu (HW). The averages of raw F_{ST} values were 0.15, 0.11, 0.16 from the AW, AH and HW comparisons, respectively. This showed that Angus and Hereford are genetically more similar than either is to Wagyu.

The distribution of genome-wide F_{ST} values for the 3 pairwise comparisons are shown in Figure 2. In total, 14 candidate regions on *Bos taurus* autosomes (BTA) 2:6, 8, 12, 13, 20, and 24 were found (Table 1). Here, we only focus on candidate regions that overlapped with previously reported regions in the literature from beef cattle QTL and association studies.

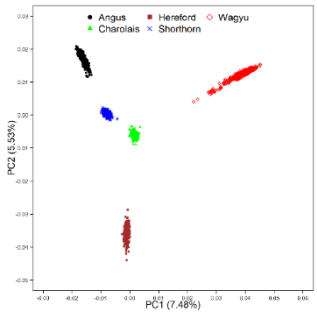

Figure 1. Plot of Principal Component 1 (PC1) vs. PC2 for 5 Australian beef cattle breeds

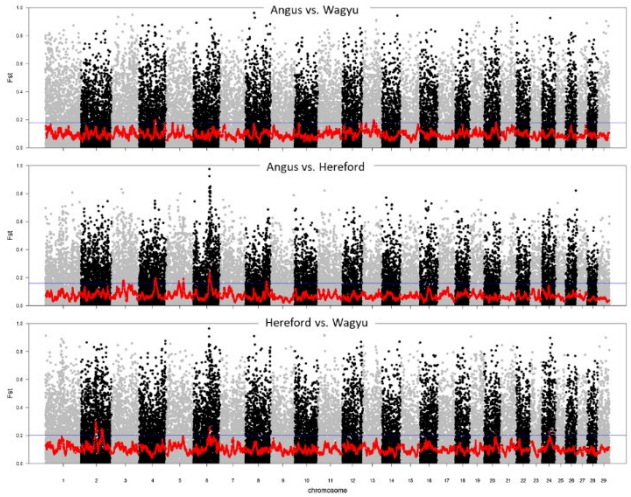

Figure 2. The Manhattan plot of genome-wide Fixation Index (F_{ST}) values. The black and grey dots show the raw F_{ST} values and the red line shows the smoothed F_{ST} values

Table 1. Candidate regions for selection

	Candidate regions*					
AW	**4**:69.96:74:30	**12**:79.81:79:87	**13**:46.24:50.91	**20**:51.45:82:74		
AH	**3**:49.54:54.27	**4**:67.15:78.32	**5**:55.99:56.80	**5**:76.29:76.92	**6**:67.32:78.89	**8**:90.69:95:74
HW	**2**:62.71:69:75	**2**:89.65:98.45	**6**:71.42:72.94	**24**:33.36:34.58		

*Chromosome:Start(Mbp):End(Mbp);

Several candidate regions found in this study have been previously reported for different economically important traits in beef cattle. One candidate region on BTA 4 overlapped between AW and AH and another candidate region on BTA 6 overlapped between AH and HW comparisons. The candidate regions on BTA 4 have been reported to contain QTLs for feed intake (Lu *et al.* 2013) and body weight (Seabury *et al.* 2017) traits in Angus and Hereford beef breeds. The candidate regions on BTA 6 intersected with regions reported to be associated with meat quality (Mateescu *et al.* 2017) and body weight (Lu *et al.* 2013) traits in Angus cattle. The candidate regions on BTA 2 from the HW comparison have been found by Snelling *et al.* (2010) to harbour variations affecting body weight in a crossbred population of different beef breeds including Hereford. A candidate region on BTA 5 between 55.99 to 56.80 Mbp from the AH comparison was also found that comprises several important genes, e.g. *INHBC*, *INHBE* and *PTGES3*, that are involved in growth

and metabolism in humans. Another candidate region on BTA 8 found in this study has been associated with feed intake (Rolf *et al.* 2012) and intramuscular fat content (Bolormaa *et al.* 2011) in Angus and Hereford cattle breeds. Mateescu *et al.* (2017) performed a genome-wide association study for meat quality traits in Angus cattle and found significant associations within the candidate region on BTA 13 found here from the AW comparison.

The candidate regions in Table 1 together encompassed 40, 197 and 91 cattle genes from the AW, AH and HW comparisons, respectively. The functional annotation clustering of the identified genes resulted in 3, 25 and 15 annotation clusters from the AW, AH and HW comparison of which only 7 clusters from the AH comparison and 2 clusters from the HW comparison were significantly enriched (enrichment score ≥ 1.3). These enriched annotation terms are associated with some biological functions e.g. embryonic skeletal system morphogenesis and protein functional domains e.g. Insulin-like growth factor-binding proteins.

CONCLUSIONS

Genome-wide screening of F_{ST} patterns provides a straightforward method to identify genomic regions under selection. Although the results here need to be validated, several candidate regions were found that may be involved in genetic differentiation between Angus, Hereford and Wagyu cattle and could explain phenotypic differences among these breeds. The candidate regions found in this study largely overlap with previously reported regions for economically important traits in beef cattle and might be useful to be incorporated in future genomic selection of Australian beef cattle.

REFERENCES

Bolormaa S., Neto L.R., Zhang Y.D., Bunch R.J., Harrison B.E., Goddard M.E. and Barendse W. (2011) *J. Anim. Sci.* **89:** 2297.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. and Lee J.J. (2015) *GigaScience.* **4:** s13742.

Durinck S., Spellman P., Birney E. and Huber W. (2009). *Nat. Protoc.* **4:** 1184.

Edriss V., Guldbrandtsen B., Lund M.S. and Su G. (2013) *J. Anim. Breed. Genet.* **130:**128.

Haerdle, W. and Steiger, W. (1995). *J. R. Stat. Soc. Ser. C. Appl. Stat.* **44**: 258.

Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009). *Nat. Protoc.* **4:** 44.

Lu D., Miller S., Sargolzaei M., Kelly M., Vander Voort G., Caldwell T., Wang Z., Plastow G. and Moore S. (2013) *J. Anim. Sci.* **91:** 3612.

Mateescu R.G., Garrick D.J. and Reecy J.M. (2017) *Front. Genet.* **8:** 171.

Rolf M.M., Taylor J.F Schnabel R.D McKay S.D McClure M.C Northcutt S.L., Kerley M.S. and Weaber R.L. (2012) *Anim. Genet.* **43:** 367.

Seabury C.M., Oldeschulte D.L., Saatchi M., Beever J.E., Decker J.E., Halley Y.A., Bhattarai E.K., Molaei M., Freetly H.C., Hansen S.L., Yampara-Iquise H., Johnson K.A., Kerley M.S., Kim J., Loy D.D., Marques E., Neibergs H.L., Schnabel R.D., Shike D.W., Spangler M.L., Weaber R.L., Garrick D.J. and Taylor J.F. (2017) *BMC Genomics*. **18:** 386.

Snelling W.M., Allan M.F., Keele J.W., Kuehn L.A., McDaneld T., Smith T.P.L., Sonstegard T.S., Thallman R.M. and Bennett G.L. (2010) *J. Anim. Sci.* **88:** 837.

Turner S.D. (2014) *J. Open Source Software.* **3:** 731.

Walmsley B.J., Donoghue K.A., Johnston D.J., Clark S.A., Siddell J.P., Walkom S.F., Granleese T. and Arthur PF. (2021) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **24:** 423.

Weir, B.S. and Cockerham, C.C. (1984). *Evolution.* **38**: 1358.

Yang, J., Lee S.H., Goddard M.E. and Visscher P.M. (2011) *Am. J. Hum. Genet.* **88:** 76.