637. A genome-wide scan for signatures of selection in Hanwoo and Angus cattle using whole-genome sequence data

H.J. Kim, S. de las Heras-Saldana, S. Clark and J.H.J. van der Werf*

School of Environmental & Rural Science, University of New England, Armidale, NSW, 2351, Australia; jvanderw@une.edu.au

Abstract

Signature of selection in livestock is a method used to identify genomic regions and candidate genes related to economic traits when phenotypic information is unknown. This study used 406 beef cattle (203 Hanwoo and 203 Angus) with a whole-genome sequence to find selection signatures. As a result, we identified 298 significant genomic regions within Hanwoo, including the candidate genes such as *LPL* related to the lipid metabolism and 33 genomic regions within Angus with *HK1* and *ACTC1* genes linked with glucose metabolism and muscle formation. From the between populations results, 36 significant genomic regions were detected, including the *CCKBR* gene associated with the feed efficiency. This study can assist in understanding the history of these breeds and identifying the genomic regions under selection for breeding programs in beef cattle.

Introduction

A signature of selection refers to genomic regions selected under natural or artificial selection (Qanbari and Simianer, 2014). The process of selection signature provides opportunities to find target selected genomic regions and genes related to economic traits, and understand the history of populations (Qanbari and Simianer, 2014). Many methods have been proposed for detecting signatures of selection. A common method is the extended haplotype homozygosity (EHH) method, which assesses the probabilities that two randomly chosen chromosomes carry a core haplotype homozygous for a region from the core SNP to a distant locus (Sabeti *et al.*, 2002). The use of haplotypes for signatures of selection studies has shown a higher potential to detect selection than other methods (Ma *et al.*, 2015). When searching for selection signatures, it is also important to select breeds with divergent attributes such as high meat yield in Angus and good meat quality in Hanwoo. It is then hypothesised that such breeds have been subjected to different breeding objectives and environments (Taye *et al.*, 2018). This study aimed to identify the signature of positive selection in Hanwoo and Angus cattle using whole-genome sequence data.

Materials & methods

A total of 406 animals with whole-genome sequence data, including 203 Hanwoo and 203 Angus, were used in this study. A quality control assessment of the genotypes in PLINK v.1.9 was applied to remove the SNPs with a genotype call rate of less than 95%, less than 1% minor allele frequency, and a *P*-value for Hardy Weinberg equilibrium <0.00001. After the quality control step, we extracted the SNPs that overlapped between both data sets resulting in the use of 7,868,872 SNPs from *Bos Taurus* autosome (BTA). All individuals were phased using Eagle v.2.4 to infer haplotypes in each population. The extended haplotype homozygosity (EHH) method was used with the R packages '*rehh*'. In addition, the integrated haplotype score (iHS) for within a population (Voight *et al.*, 2006) and the Rsb statistic (Tang *et al.*, 2007) for between populations assessment were used. The values of iHS and Rsb were standardized using a mean of 0 and a standard deviation of 1 for allele frequency at the core SNP and transferred from each value to the *P*-value using the Gaussian cumulative distribution function. The significance threshold for each analysis was used for iHS ($-\log_{10}$*P*-value (0.0001)>4) and for Rsb ($-\log_{10}$*P*-value(0.05)>5.48). The significant genomic regions were defined within 1 Mb of significant SNPs, and extended until there were no significant SNPs. We used the Ensembl database with the ARS-UCD1.2 bovine reference genome

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_637, © H.J. Kim *et al.* 2022

to identify the candidate genes in significant genomic regions. These genes were compared with those previously identified in the Cattle QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/BT/search).

Results

We constructed Manhattan plots with the *P*-value of iHS scores for each population and Rsb between populations (Figure 1). As a result, 298 significant genomic regions (-$\log_{10}$(*P*-value)>4) were detected for Hanwoo, but only thirty-three regions were identified in Angus. The most significant value of iHS was located on BTA6 (iHS=6.95) in Hanwoo and BTA10 (iHS=4.87). For the signature of selection results of the between population analysis using Rsb, 36 significant regions were detected and the most significant value of Rsb was -6.55 on BTA12.

Nine candidate genes surrounding the top 5 significant regions within the Hanwoo population were detected on BTA8 and BTA16. For Angus, there were 50 candidate genes detected in the top 5 significant regions on BTAs 6, 10, 20, 28 and 29. From the between population results, 32 candidate genes were identified in the top 5 significant genomic regions. Interestingly, the region on BTA28 was significant and shared between populations (Rsb) and the iHS results in Angus (Table 1).

Discussion

The EHH values were used to detect signatures of selection in each population. iHS, the EHH method for within population analysis, compares the ancestral allele and derived allele, while Rsb compares between populations. The iHS method found 298 and 33 significant genomic regions for Hanwoo and Angus, respectively. Genes related to the immune system and intramuscular fat were previously reported in Hanwoo (Lim *et al.*, 2016) on BTAs 2, 3, 9, 10, 11, 12, 22 and 26. But genes involved in body size, body structure and meat quality were identified for Angus on BTAs 2, 3, 4, 14, 15 and 16 (Paim *et al.*, 2020). Some significant genomic regions that overlapped with previous studies were on BTA2 and BTA11 in Hanwoo and on BTA14 in Angus.

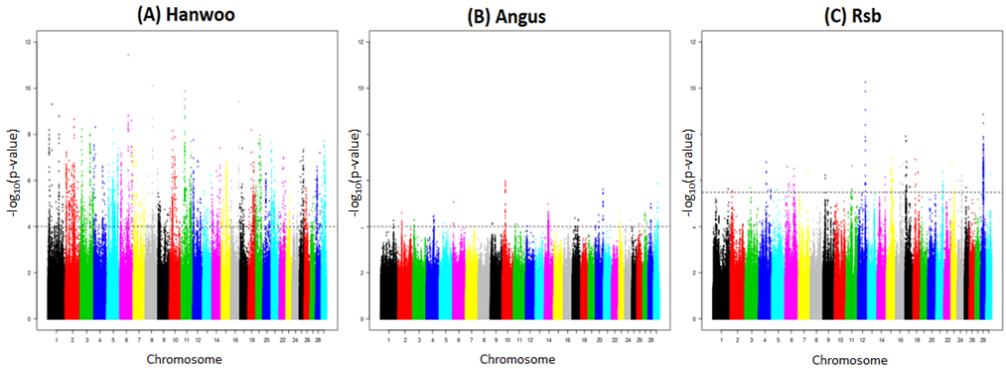

Figure 1. The Manhattan plot of -$\log_{10}$(*P*-value) of (A) iHS in Hanwoo, (B) iHS in Angus, and (C) Rsb between Hanwoo and Angus.

Table 1. The top 5 significant genomic regions and associated genes.[1]

Method	Population	Chr	Regions (Mb)	n SNP	Genes
iHS	Hanwoo	6	75.2-76.2	18	-
		8	66.1-67.1	34	*LPL*
		11	34.2-35.2	126	-
		16	75.2-76.2	12	*PLXNA2, CD34, CD46, ASPM, ZBTB41, F13B, CRB1*
		1	36.5-37.5	3	*EPHA3*
	Angus	10	29.9-30.9	27	*GREM1, SCG5, GJD2, ACTC1, AQR, ZN770, DPH6*
		29	38-39.5	13	*PAG7, PAG15, PAG4, PAG14, PAG16, PAG20, PAG21, PAG1, PAG19, PAG17, MGC157408, MGC157405, PAG9, PAG3, PAG6, PAG11*
		20	70.4-71.5	14	*IRX4, NDUF56, LPCAT1, SLC6A3, CLPTM1L, TERT, SLC6A19, SLC6A18, NKD2, TRIP13, BRD9, TPPP, CEP72*
		6	7.84-8.87	3	*TRAM1L1*
		28	24.9-25.9	12	*TET1, CCAR1, STOX1, DDX50, DDX21, KIFBP, SRGN, VPS26A, SUPV3L1, HKDC1, HK1, TACR2, TSPAN15*
Rsb	Between	12	69.9-70.9	18	-
		28	24.8-25.9	240	*SLC25A16, TET1, CCAR1, STOX1, DDX50, DDX21, KIFBP, SRGN, VPS26A, SUPV3L1, HKDC1, HK1, TACR2, TSPAN15*
		17	6.6-7.9	8	*SH3D19, RPS3A, LRBA, MAB21L2, DCLK2*
		17	13.5-14.7	16	*HHIP, GYPA, GYPB, SMARCA5, GAB1, USP38*
		15	46.7-49.1	33	*OR52fam, CAVIN3, CCKBR, CNGA4, C15H11orf42, HBE2, TRIM34*

[1] Chr = chromosome; Mb = mega bases; n SNP = The number of SNPs in the region.

The results of iHS in Hanwoo were stronger than Angus. We assumed that Hanwoo might have had less selection pressure than Angus because the effective population size of Angus decreased faster than Hanwoo in this study. iHS is a more powerful method to detect intermediate selection sweeps than complete selection signatures. A population with weak selection pressure can have a higher number of intermediate allele frequencies than a population with strong selection pressure due to relatively taking more time for complete sweeps under a weak selection pressure population (Lopez *et al.*, 2019). This might be the reason why more significant regions were found with iHS in Hanwoo. In contrast with iHS, Rsb has the power to detect complete selective sweeps (Tang *et al.*, 2007). Most significant genomic regions from Rsb were related to Angus since this breed shows more evidence of strong selection. The nine candidate genes located close to the top 5 significant regions for the within population (iHS) assessment in Hanwoo were *LPL* and *ZBTB41*, which have been associated with the intramuscular fat in Qunchaun cattle (Oh *et al.*, 2013), and residual feed intake in Holstein (Zhou *et al.*, 2018), respectively. For Angus, the *HK1* and *ACTC1* genes were linked to glucose metabolism (Ebara *et al.*, 2013) and muscle formation (Qanbari *et al.*, 2011). These results confirm that Hanwoo has been mainly selected for meat quality traits (lipid metabolism), while Angus has been selected for growth and meat yield. For Rsb, 32 candidate genes were detected from which the *LRBA* gene was associated with the kinase A related to the immune effector of molecules (Mapholi *et al.*, 2016), and the *CCKBR* gene was associated with feed efficiency in cattle (Abo-Ismail *et al.*, 2013). These results can be helpful to confirm the evidence of selection and pre-selecting these SNPs could help to increase the genomic prediction accuracy (Ye *et al.*, 2020) for economically important traits in beef cattle breeding programs.

References

Abo-Ismail M.K., Kelly M.J., Squires E.J., Swanson K.C., Bauck S *et al.* (2013) J Anim Sci 2512-2529. https://doi.org/10.2527/jas2012-5756

Ebara F., Inada S., Morikawa M., Asaoka S.H., Isozaki Y *et al.* (2013) J Anim Physiol Anim Nutr (Berl) 97(4), 684-693. https://doi.org/10.1111/j.1439-0396.2012.01310.x

Iyengar V.K., Reeve H.K., and Eisner T. (2002) Nature 419(6909), 830-832. https://doi.org/10.1038/nature01027

Lim D., Choi B.H., Cho Y.M., Chai H.H., Jang G.W *et al.* (2016) BMB Rep 49(9), 514-519. https://doi.org/10.5483/bmbrep.2016.49.9.074

Lopez M.E., Linderoth T., Norris A., Lhorente J.P., Neira R *et al.* (2019) Front Genet, 10, 901. https://doi.org/10.3389/fgene.2019.00901

Ma Y., Ding X., Qanbari S., Weigend S., Zhang Q *et al.* (2015) Heredity (Edinb) 115(5), 426-436. https://doi.org/10.1038/hdy.2015.42

Mapholi N.O., Maiwashe A., Matika O., Riggio V., Bishop S.C *et al.* (2016) Ticks Tick Borne Dis 7(3), 487-497. https://doi.org/10.1016/j.ttbdis.2016.02.005

Oh D., La B., Lee Y., Byun Y., Lee J. *et al.* (2013) Mol Biol Rep 40(4), 3155-3163. https://doi.org/10.1007/s11033-012-2389-y

Paim T.D.P., Hay E.H.A., Wilson C., Thomas M.G., Kuehn L.A *et al.* (2020) Front Genet 11, 710. https://doi.org/10.3389/fgene.2020.00710

Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S *et al.* (2011) BMC Genomics 12:318. https://doi.org/10.1186/1471-2164-12-318

Qanbari S., and Simianer H. (2014) Livestock Science 166, 133-143.

Shi Y., Zhao P., Dang Y., Li S., Luo L *et al.* (2021) Biol Reprod 105(2), 359-370. https://doi.org/10.1093/biolre/ioab081

Strillacci M.G., Moradi-Shahrbabak H., Davoudi P., Ghoreishifar S.M., Mokhber M *et al.* (2021) BMC Genomics 22(1), 305. https://doi.org/10.1186/s12864-021-07604-3

Tang K., Thornton K.R., and Stoneking M. (2007) PLoS Biol 5(7), e171. https://doi.org/10.1371/journal.pbio.0050171

Voight B.F., Kudaravalli S., Wen X., and Pritchard J.K. (2006) PLoS Biol 4(3), e72. https://doi.org/10.1371/journal.pbio.0040072

Ye S., Song H., Ding X., Zhang Z., and Li Z. (2020) Animal 14(8):1555-64. https://doi.org/10.1017/S1751731120000506

Zhou Y., Connor E.E., Wiggans G.R., Lu Y., Tempelman R.J *et al.* (2018) BMC Genomics 19:314. https://doi.org/10.1186/s12864-018-4699-5