

Selection of genomic regions and genes associated with adaptation and fertility traits in two Colombian creole cattle breeds

C. De León¹, R. Martínez², J.F. Rocha² and A.E. Darghan³

¹ Servicio Nacional de Aprendizaje SENA, Colombia

² Corporación Colombiana de Investigación Agropecuaria-Agrosavia
Mosquera, Cundinamarca, Colombia

³ Universidad Nacional de Colombia, Bogotá, Colombia

Corresponding author: C. De León

E-mail: crdeg@unal.edu.co

Genet. Mol. Res. 20 (3): gmr18882

Received April 21, 2021

Accepted June 25, 2021

Published July 31, 2021

DOI <http://dx.doi.org/10.4238/gmr18882>

ABSTRACT. Colombian creole breeds Blanco Orejinegro (BON) and Sanmartinero (SM) are widely used as pure breeds and in crossbreeding programs due to their excellent performance in the double ability of milk and meat production. We examined genomic regions and genes that from generation to generation have been selected by positive natural selection in favor of the adaptability and reproductive performance of these two Colombian creole breeds. Natural selection of genomics region and genes is one of the main evolutive changes resulting in phenotypic adaptations. The selected genomic regions can be detected by comparing differences in regional linkage disequilibrium (LD) between cattle populations with potential adaptations for economic traits. In this study, we used 58,868 single nucleotide polymorphisms (SNPs) from BON and 57482 SNPs from SM, using genotyping data from 1262 BON and 742 SM animals to estimate the variation of genome-wide LD between populations using the VARLD program. The top 0.1 and 0.01th percentiles of standardized VarLD scores were used as a criterion for all comparisons. A total of 10 selection signatures on chromosomes 3, 5, 11, 15, 18, 21, 22, 23, 25 and 29 were identified in all populations. These signatures overlapped with quantitative trait

loci for adaptability and reproductive performance in both breeds. Within the signature located between 23,903,882 and 23,955,588 bp on chromosome 18, we identified the *SLC6A2* gene involved in the response to high stress. Also, we identified the genes *CTDSP2*, *CES1*, *CFAP161*, *CLEC14A*, *HIPK1*, *RBM4*, *SSTR* involved in the expressions of economic traits (meat and milk production), *KDMID*, *OLFML3* genes involved in reproductive traits (age at first calving and calving interval), and *ATP23*, *LRRTM1*, *SLC6A2*, *DEK*, *SYT6*, *KDMID* genes involved in cellular stress response and response to important environmental changes such as high temperature. We conclude that these genomic regions seem to point toward a recent selection in BON and SM populations. These regions can be used in selection and conservation programs.

Key words: Creole breeds; SNPs; Adaptability indexes; LD variation

INTRODUCTION

The Blanco Orejinegro (BON) and Sanmartinero (SM) Colombian creole cattle breeds are included in the national genetic conservation and breeding program because of their excellent adaptation and reproductive performance in tropical production systems. This has led to genetic characterization studies for applying results obtained in this program. BON and SM breeds have been genetically characterized by Martínez et al. (2012); however, this information has had limited impact because of a lack of studies assessing in depth the variation and genetic differences of these two breeds. These differences can be measured through determination of selection signatures, which are genomic regions that may be fixed in individuals within a population as a result of artificial or natural selection for reasons such as adaptability or productivity (Msalya et al., 2017).

Linkage disequilibrium (LD; non-random association of alleles of different loci located in the same chromosome) can be used to detect positive selection (Montgomery, 2008). Positive selection (a type of natural selection that favors a single allele and thus the allele frequency of a population continually goes in one direction) affects the patterns of LD around the site of a beneficial allele increasing the LD among the alleles due to selection on the nearby favorable allele (Alachiotis and Pavlidis, 2016). The frequency of favorable alleles increases rapidly through adaptation and selection when a new beneficial mutation appears in a population. Therefore, identifying these selection signatures may help in the exploration of genotype–phenotype relationships to improve economically important traits (Biswas and Akey, 2006).

Selection is responsible for maintaining from generation-to-generation favorable alleles increasing its frequency until reaching fixation (Pérez O'Brien et al., 2014). The strong selection pressure increases frequency of the favorable mutation until completely replacing the ancestral allele by selective sweep-SW. SW perform a certain pattern of variation around the selected position, different from that expected under neutrality (Walsh and Lynch, 2018).

Several methodologies have been proposed to measure patterns of variability of regions throughout the genome, such as those proposed by Sabeti et al. (2002, 2006, 2007),

extended haplotype homozygosity (EHH) and Integrated haplotype score (iHS) proposed by Voight et al. (2006). However, these methodologies are limited because the SNPs statistically significant can be biologically inactive, although they are probably in LD with some unknown functional polymorphism. To obtain greater precision in the results, Teo et al. (2009) propose to explore LD variation (varLD) of these regions because, according to them, this variation is biologically significant. The varLD methodology basically consists of exploring the differences of the LD patterns between chromosomal regions of animal or human populations. This methodological strategy becomes an excellent tool to build fine genetic mapping of candidate regions (Signature of selection) associated to diseases, reproduction or production and adaptation traits (Teo et al., 2009).

Therefore, the aim of this study was to look for signatures of selection or genomic regions that increased in frequency and were fixed in the population, as well genes annotated in them, since these processes of natural or artificial selection, improve the adaptative and productive ability in the population, the results obtained can be applied in programs of conservation and genetic improvement of these two Colombian creole cattle breeds and for mating with another breeds.

MATERIAL AND METHODS

Animals and study location

Animals come from two Colombian creole cattle breeds, Blanco Orejinegro (BON) and Sanmartinero (SM), obtained from Germplasm Bank of Corporación Colombiana de Investigación Agropecuaria (Agrosavia, 2019). BON animals are maintained in the experimental station El Nus in San Roque, department of Antioquia, in a mountainous area in the Andean region with altitudes ranging from 800 to 1,800 meters. Topography is undulating with annual average temperatures between 18 and 24 °C and a bimodal rainfall regime. SM animals are located in the experimental station La Libertad in Villavicencio, department of Meta, in Colombian Orinoquia on flat lands with heights of 336 m.a.s.l. in the humid tropical forest zone. There is a rainy season from April to November and a dry season from December to March, with annual average temperatures that range between 27 and 41°C and relative humidity from 36 to 87%. Both BON and SM populations were maintained under semi-extensive grazing conditions.

Data source

DNA samples extracted from blood and / or semen were genotyped at the C.I. Genomics Laboratory Agrosavia Tibaitatá following the Infinium HD Assay protocol (Illumina Inc San Diego, CA, USA, 2016), at the end of the process to obtain the genotypes each chip was scanned in a HiScan® device and the generated data was analyzed with the GenomeStudio software (Illumina, 2016), which assigns the genotypes of each SNP based on signal differences and color intensity, and allows a preliminary edition of the data in terms of genotyping quality (Call Rate and Gentrain). This procedure generated 58868 SNPs (Single Nucleotide Polymorphism) in the BON breed and 57482 in the SM breed, these SNPs come from 1,262 genotyped animals of the BON breed and 742 from the SM breed, Table 1.

Table 1. Genotyped animals from Blanco Orejinegro (BON) and Sanmartinero (SM) breeds.

Breed	20KGeneSeek	BovineSNP50K	IlluminaSNP7K	Total
BLANCO OREJINEGRO	490	643	129	1262
SANMARTINERO	418	198	126	742

1262 and 742 BON and SM animal respectively was Genotyped. Different types of chips were used for all animals, 7K (7000 SNPs), 20K (20000 SNPs) and 50K (50000 SNPs). Source: *Agrosavia Genomics Laboratory*.

Regarding BON breed of a total of 1262 genotyped animals, 129 animals were genotyped using the IlluminaSNP7K low density chip, 490 animals with the 20KGeneSeek chip and 643 animals with the BovineSNP50K medium density chip. Regarding SM breed, of a total of 742 genotyped animals, 126 were genotyped with low density chip IlluminaSNP7K, 418 animals with 20KGeneSeek chip and 198 with the medium density chip BovineSNP50K. SNPs and their genotypes from the 7K and 20K chips were imputed to the 50K chip using the Fimpute program (Sargolzaei et al., 2014). After imputation, quality control (Quality Control-QC) was performed on 58868 available SNPs to BON breed and 57482 to SM breed, using PLINK programs (Purcell et al., 2007). SNPs were excluded, if the call rate was less than 90%, if they presented an extreme deviation from Hardy-Weinberg equilibrium ($P < 0.01$) (for example, SNP in autosomal chromosomes presenting both homozygous genotypes, but without heterozygotes), the same for SNP with unknown genomic position, or if they were located in sex chromosomes, also if they were monomorphic or if frequency of less common allele (MAF) was below 0.03. Likewise, animals that had a Call Rate lower than 90% or with pedigree errors or Mendelian conflicts between relatives were eliminated. From result of this process, were obtained 51283 effective polymorphic SNPs located in autosomal chromosomes for BON breed and 49009 SNPs for SM breed.

Statistical analysis

The statistical analysis is based on varLD scores between genomic regions of BON and SM populations. The analysis was carried out in several phases. First, LD was initially quantified using correlation equation between alleles of two SNPs, then varLD scores between all regions, is estimated, using arrays matrix obtained from SNP information using the program R. Standardization of varLD scores was made taking scores of the percentiles 99.9th and 99.99th (Pérez O'Brien et al., 2014). For estimating percentiles in genomic regions, taken into account means and variances obtained from SNPs in whole genome like in Teo et al. (2009). Percentiles were calculated and plotted in R program.

LD quantification

r^2 assess strength of genetic correlation between alleles of two SNPs. r^2 quantifies how much LD increases or decreases between two SNPs reflecting, thus the direction of correlation between two SNPs, so r^2 is the most appropriate measure to compare LD (Teo et al., 2009). r^2 for two biallelic SNPs with alleles (A, a) and (B, b), is defined as follows.

$$r^2 = \frac{(P_{AB} - P_A P_B)^2}{P_A P_a P_B P_b} - (-1)^{I(P_{AB} < P_A P_B)} \quad (\text{Eq. 1})$$

where P_{AB} is the frequency of the AB haplotype; P_A , P_a , P_B and P_b , are the respective allelic frequencies; the term $I(P_{AB} < P_A P_B)$, denotes a function indicator that takes a value of one when $P_{AB} < P_A P_B$ and zero otherwise.

varLD scores on regions, regional varLD

varLD scores are obtained from construction of two symmetric correlations matrices, M_1 and M_2 one for each population, so from population k it's obtained $M_k = T_k \Delta_k T_k^T$ with columns T_k denoting eigenvectors of M_k . Δ_k is a diagonal matrix of categorized values of M_k in descending order. varLD score is obtained from subtraction of two diagonals $|\Delta_1 - \Delta_2|$, this score provides a measure of dissimilarity degree between correlation matrices M_1 and M_2 , this value will be used to quantify regional LD differences in the two populations. Matrices elements M_1 and M_2 are formed by consecutive overlapped sliding windows of L SNPs where each consecutive window is obtained by moving the existing window in the direction of the forward chain for a SNP. For a particular window of L SNPs, (i, j) entry in each matrix represents the r^2 between SNP i and SNP j (Teo et al., 2009). Therefore, each of these matrices effectively represents a correlation matrix between L SNPs in the respective population (Zaykin et al., 2006). All these calculations were done in R.

Statistical significance for each window of L SNPs

The statistical significance for each window of L SNP was obtained from non-deterministic Monte Carlo procedure for evaluating evidence against the null hypothesis of no differences in LD structure in each region of L SNP (Krzanowski, 1993). Significance by Monte Carlo procedure is defined as $(M + 1) / (Niter + 1)$ where M is the number of varLD scores obtained from resampling scheme and Niter is the combinations number. Combinations result from resampling of r^2 values of both populations. This results in combined data with replacement, produce two populations of equal size, from this, it can calculate varLD value and this value is applied in the Monte Carlo significance test $(M + 1) / (Niter + 1)j$ (Teo et al., 2009).

Standardized varLD scores

Because the statistical significance can be affected by the density of the SNPs, it is advisable to standardize varLD score, for this the 99.99th and 99.9th percentiles are taken as reference (Pérez O'Brien et al., 2014). Once the varLD scores are calculated, they are standardized using the formula $S_i' = (s_i - E(s)) / \sqrt{\text{var}(s)}$, in R, where $E(s)$ and $\text{var}(s)$ are the means and variances of values of the whole genome, to center the distribution of the scores around an average of zero and a standard deviation of one, which helps to avoid biases in the VarLD scores due to differences in size of the windows in terms of base pairs (bp) and the populations that are compared have different LD background levels. The software uses sliding windows and windows of L SNPs are applied following Teo et al. (2009). A window was marked as an assumed Signatures of selection (SS region) if the score S_i' was greater or equal to the score in the 99.99th percentile and the 99.9th percentile of all scores across the

genome. The central position of the first window in an identified SS was taken as the starting point of a signal, and the final position was half of the final window in the SS.

Determination of signatures of selection

To assess selective sweep in a particular region is relatively easy when the levels of r^2 differ greatly, as in this study, the standardized varLD score, S_i' , was high, because of there are higher LD levels in that particular region. The differences of the LD value in the SS regions identified, can be seen in the standardized varLD graphs using R. Only the graphs corresponding to the regions explored in detail in this study are presented.

Genes Identification

After determination of candidate SS regions, details such as the chromosome and start position and end position, windows included in the peak were taken into account. Then, SS regions were classified by the position on the chromosome in bp. To identify genes possibly associated with the SS regions, regions position in bp were listed in the Ensembl tool (Ensembl Biomart, 2019) to visualize genes in common position with SS regions.

Functional analysis of genes

Functional analysis of the mapped genes was carried out through the website UniProt, (2019) and GeneCards® (2019), which were used to verify genes functional information. When no information was available for *Bos taurus* genes, annotations of orthologous genes were used in humans, rats or mice to proceed with the functional analysis, which allowed to define a list of genes located within or surrounding regions of interest.

RESULTS

Strong SS were limited to narrow regions of the genome. The most distinctive peaks were observed on chromosomes 3, 5, 11, 15, 18, 23, 25 and 29 (Figure 1) that showed the differences in the percentile distributions with scores highest 0.01 and 0.1 percentiles, Table 2.

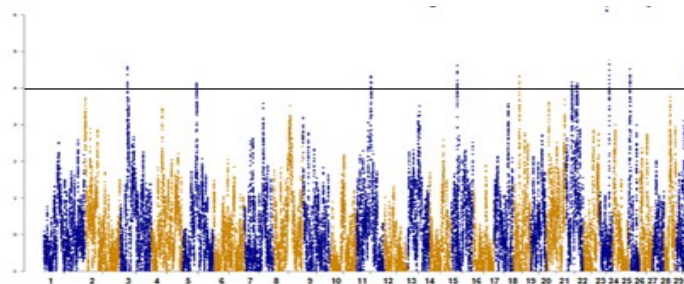


Figure 1. Linkage disequilibrium variation analysis in the entire genome of Blanco Orejinegro and Sanmartinero Breeds. Note the highest peaks corresponding to standardized varLD scores above 4.0, on chromosomes 3, 5, 11, 15, 18, 21, 22, 23, 25 and 29.

Table 2. Genomic regions in base pairs (bp) on chromosomes 3,5,11,15,18,21,23,29, Linkage disequilibrium Variations and Standardized varLD scores, on percentiles 0.01 and 0.1 and genes found in these regions.

Chr/Percentiles	Regions (bp)	varLD	Standardized varLD scores	Genes
Chr3				
P 0.01	29.224.886	21.7	4.58	
P 0.1	29.285.827	21.6	4.54	
Chr5				
P 0.01	55.085.241	20.3	4.13	<i>ATP23, CTDSP2.</i>
P 0.1	55.582.258	20.2	4.10	
Chr11				
P 0.01	55.683.407	20.9	4.33	<i>LRRM1.</i>
P 0.1	55.443.099	20.8	4.30	
Chr 15				
P 0.01	24.184.623	21.8	4.61	<i>ZW10</i> 24.284.251-
P 0.1	24.186.798	21.3	4.45	24.319.569
Chr18				
P 0.01	23.903.882	20.9	4.32	<i>SLC6A2, CES1</i>
P 0.1	23.955.588	20.8	4.30	
Chr21				
P 0.01	26.990.623	20.3	4.15	<i>CFAP161</i>
P 0.1	26.929.649	19.9	4.02	
P 0.01	48.369.157	20.2	4.12	<i>SSTR1</i>
P 0.1	48.313.773	19.9	4.01	
Chr23				
P 0.01	39.339.694	23.9	5.24	<i>DEK:</i> 39.355.962- 39.387.656
P 0.1	39.354.709	23.9	5.22	<i>KDM1B:</i> 39.389.854- 39.432.004
Chr29				
P 0.01	44.796.094	21.9	4.64	<i>SPTBN2, RBM4B</i>
P 0.1	44.707.685	21.8	4.59	

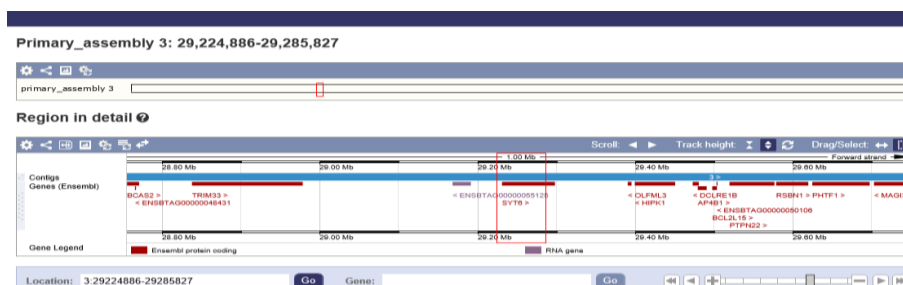


Figure 2. Genomic region (Selection Signature) identified on chromosome 3 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome. Selection signature (SS) identified in the red rectangle has annotations for the *SYT6* gene. Indicating that this SS and the SNPs within it, are highly associated to this gene. Also, very close to this SS are the *OLFM3* and *HIPK1* genes at 16 and 17.5 Kb, respectively, possibly associated to this SS.

The chromosomes BTA3, BTA5, BTA11, BTA15, BTA18, BTA21, BTA23, BTA25 and BTA29 showed higher standardized varLD score than the rest, with score ≥ 4.00 on percentiles 0.01 and 0.1 with peaks indicating signatures of selection (See Figure 1).

On BTA3, one signature of selection (SS) was identified between 29,224,886 and 29,285,827 bp with standardized varLD score between 4.54 and 4.58 in the 0.01th and 0.1th percentiles (Table 2, Figure 2, Figure 2(a)). In this region annotations for the *SYT6* gene

(Synaptotagmin 6) were found. Surrounding this SS at 10 Kb, two genes, possibly associated to this SS, were identified. The *OLFM3* gene (Olfactomedin-like protein 3) to 29,390,850-29,394,608 bp, and *HIPK1* gene (Homeodomain interacting protein kinase 1) at 29,399,611-29,449,947 bp. (Ensembl, 2019) (Table 2, Supplementary 1).

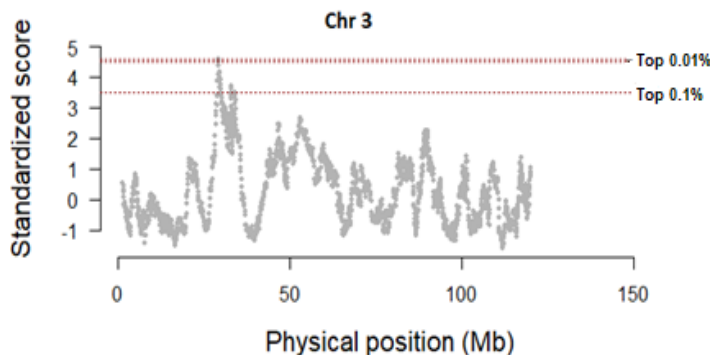


Figure 2(a) . Physical position in megabases (Mb) on chromosome 3 of the genomic region according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

On BTA5 one SS between 55,085,241 and 55,582,258 bp was identified (Table 2; Figure 3, Figure 3(a)), with annotations for *ATP23* gene (Mitochondrial inner membrane protease ATP23). Near of this region at 5.5 Kb, the *CTDSP2* gene (Carboxy-terminal domain, RNA polymerase II, polypeptide A small phosphatase 2) at 55,634,355-55,656,768 bp was located, possibly associated with this SS (Ensembl, 2019).

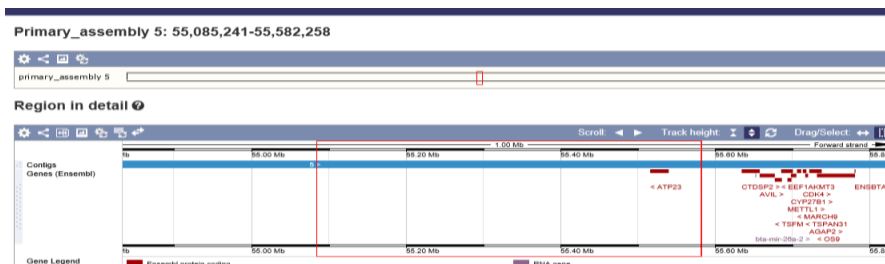


Figure 3. Genomic region (Selection Signature) identified on chromosome 5 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome. Selection signature (SS) identified in the red rectangle has annotations for the *ATP23* gene. Indicating that this SS and the SNPs within it, are highly associated to this gene. Also, very close to this SS are the *CTDSP2* gene at 5.5 Kb approximately, possibly associated to this SS.

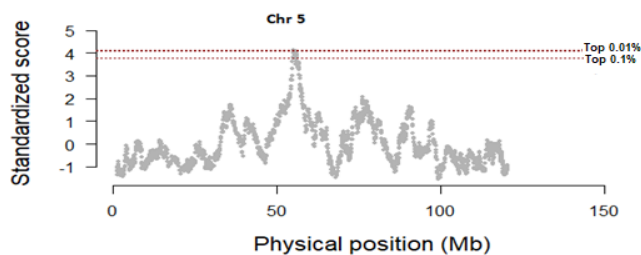


Figure 3(a). Physical position in megabases (Mb) on chromosome 5 of the genomic region according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

On BTA11, one SS between 55,443,099-55,683,407 bp was identified (Table 2, Figure 4, Figure 4(a)), no annotations for genes in this region (Table 2, Annex 1). However, at 230 Kb, the *LRRTM1* gene (Leucine-rich repeat transmembrane neuronal protein 1) at 55,211,496-55,214,132 bp. was identified, possibly associated with this SS. mechanisms.

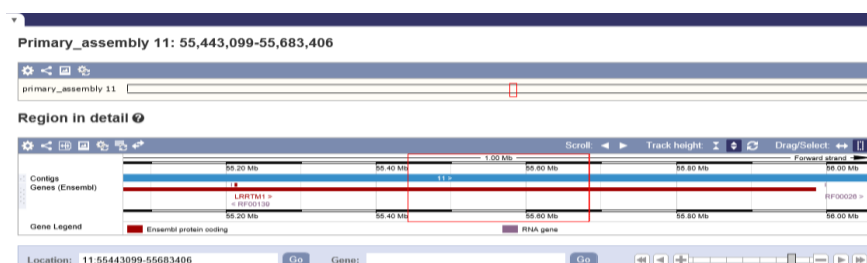


Figure 4. Genomic region (Selection Signature) identified on chromosome 11 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome. Selection signature (SS) identified in the red rectangle hasn't gene annotations. Outside of rectangle the *LRRTM1* gene is located at 230 Kb from this region, possibly associated to this SS.

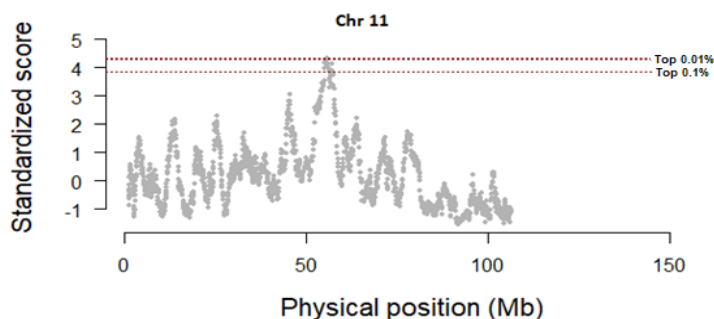


Figure 4(a). Physical position in megabases (Mb) on chromosome 11 of the genomic region according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

On chromosome 15, one SS between 24,184,623 and 24,186,798 bp was identified, no annotations for genes in this region (Table 2; Figure 5; Figure 5 (a); Annex 1). However, near this SS at 100 Kb, the *ZW10* gene (Centromere / kinetochore protein zw10) at 24,284,251-24,319,569 bp, was found (Ensembl, 2019) (Table 2; Annex 1).

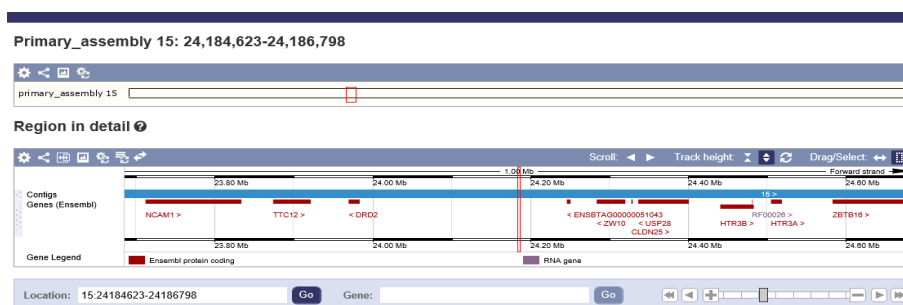


Figure 5. Genomic region (Selection Signature) identified on chromosome 15 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome. Selection signature (SS) identified in the red rectangle hasn't genes annotations. Outside of rectangle the *ZW10* gene is located at 100 Kb from this region, possibly associated to this SS.

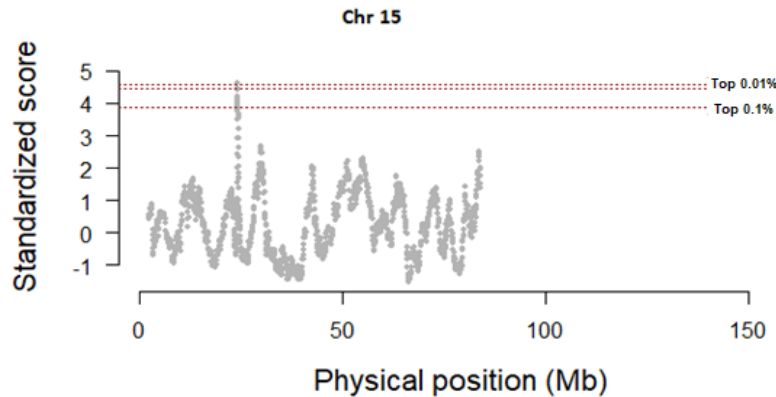


Figure 5(a). Physical position in megabase (Mb) on chromosome 15 of the genomic region according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

On BTA18, one SS between 23,903,882 and 23,955,588 bp, was identified (Table 2; Figure 6; Figure 6 (a)) with annotations for the *SLC6A2* (Transporter) gene and *CES1* (Carboxylic ester) gene. (Ensembl, 2019) (Table 2; Annex 1).

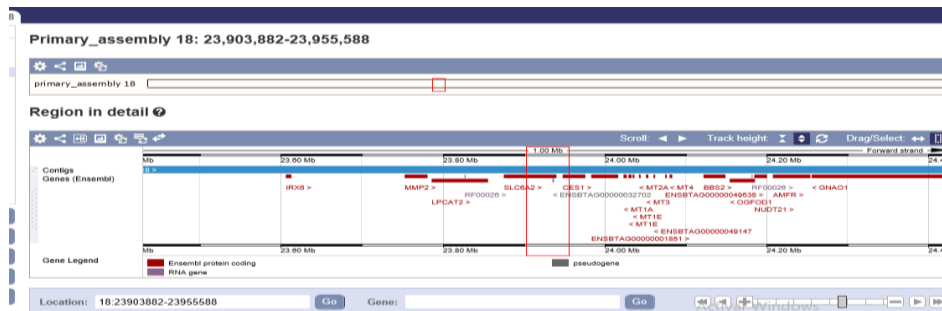


Figure 6. Genomic region (Selection Signature) identified on chromosome 18 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome Selection signature (SS) identified in the red rectangle has annotations for the *SLC6A2* and *CES1* genes. Indicating that this SS and the SNPs within it, are highly associated to these genes.

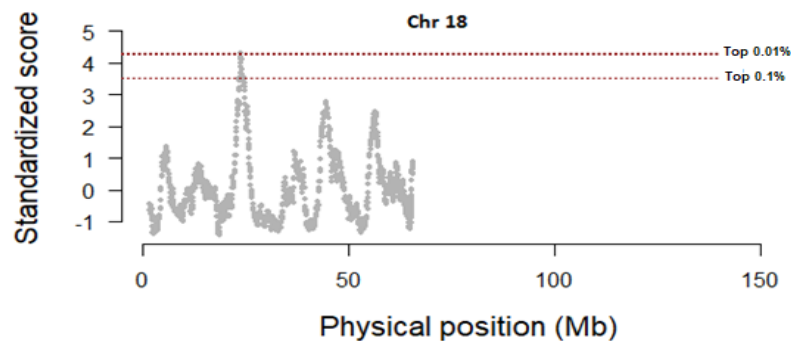


Figure 6(a). Physical position in megabases (Mb) on chromosome 18 of the genomic region according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

On BTA21, three SS were identified. The first one, between 26,929,649 and 26,990,624 bp, was located, no annotations for genes in this region. However, near this SS at 9 Kb, the *CFAP161* gene (Cilia and flagella associated protein 161) was found. The second one, between 48,313,773 and 48,369,157 bp was located, with annotations for the *SSTR1* gene (Somatostatin receptor 1). The third SS, between 48,594,114 and 48,651,030 bp was located (Table 2, Figure 7, Figure 7 (a)), no annotations for genes. However, near this SS at 200 Kb, the *CLEC14A* gene (C-type lectin domain family 14, member 1) was found. (Ensembl, 2019) (Table 2).

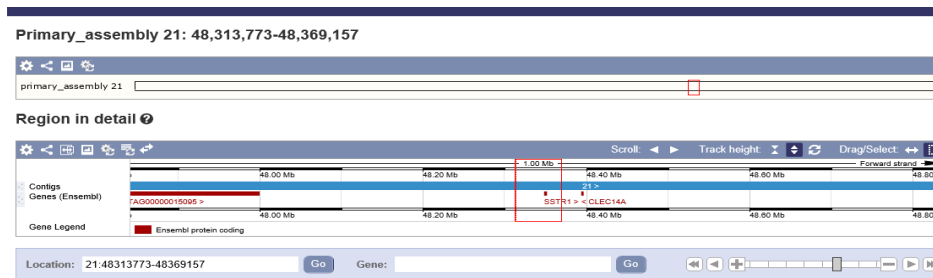


Figure 7. Genomic region (Selection Signature) identified on chromosome 21 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome Selection signature (SS) identified in the red rectangle has annotations for the *SSTR1* gene. Indicating that this SS and the SNPs within it, are highly associated to these gene.

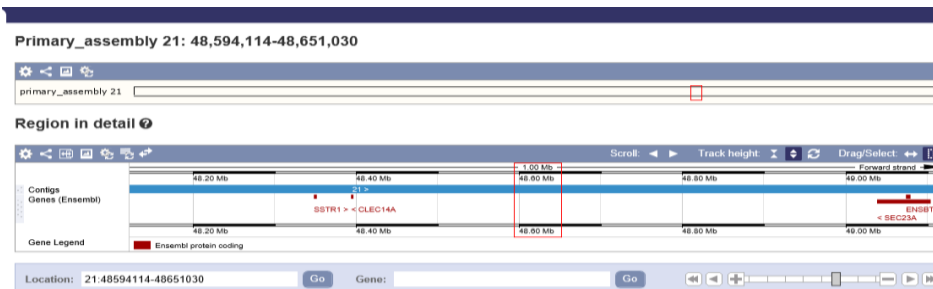


Figure 7. Genomic region (Selection Signature) identified on chromosome 21 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome Selection signature (SS) identified in the red rectangle hasn't genes annotations. Outside of rectangle the *CLEC14A* gene is located at 200 Kb from this region, possibly associated to this SS.

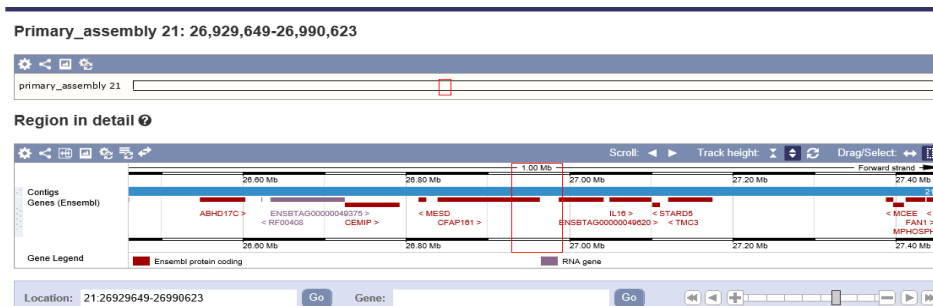


Figure 7. Genomic region (Selection Signature) identified on chromosome 21 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome Selection signature (SS) identified in the red rectangle hasn't genes annotations. Outside of rectangle the *CFAP161* gene is located at 100 Kb from this region, possibly associated to this SS.

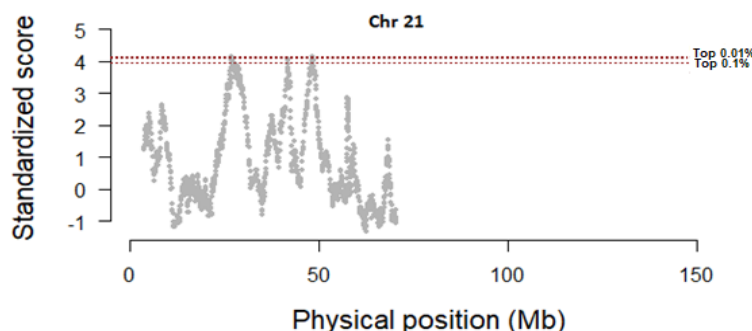


Figure 7(a). Physical position in megabases (Mb) on chromosome 21 of the 3 genomic regions according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

On BTA23, one SS between 39,339,694 and 39,354,709 bp was identified (Table 2, Figure 8, Figure 8 (a)) with standardized varLD score between 5.0 and 5.1, no annotations for genes. However, near to this region at 3.5 Kb the *DEK* gene (DEK protein) at 39,355,962-39,387,656 bp and at 9 Kb the *KDM1B* gene (Lysine demethylase 1B) at 39,389,854-39,432,004 bp were found. (Ensembl, 2019) (Table 2, Annex 1). These genes are possibly associated with this SS.

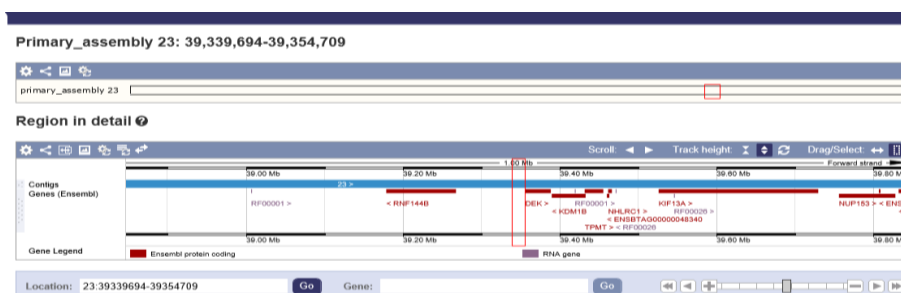


Figure 8. Genomic region (Selection Signature) identified on chromosome 23 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome Selection signature (SS) identified in the red rectangle hasn't genes annotations. Outside of rectangle the *DEK* and *KDM1B* genes are located at 3.5 and 9 Kb respectively, possibly associated to this SS.

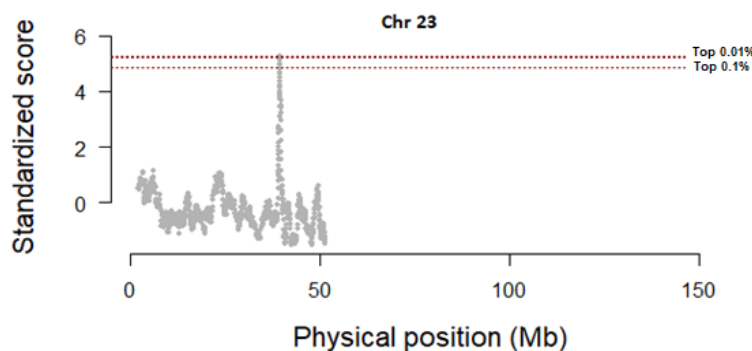


Figure 8(a). Physical position in megabases (Mb) on chromosome 23 of the genomic region according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

On BTA25, one signature of selection between 5,671,356 and 6,189,134 bp was identified (Figure 9; Figure 9(a)) no genes annotation for this region.

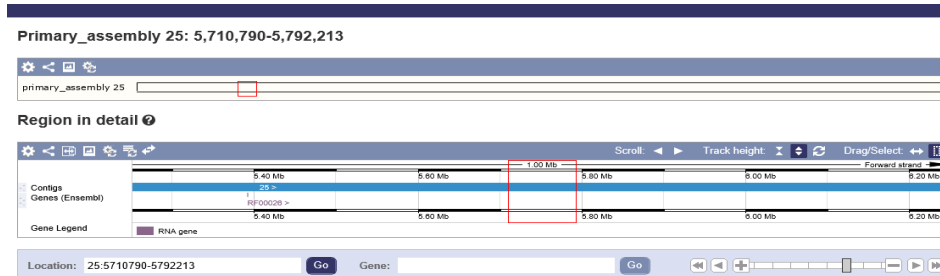


Figure 9. Genomic region (Selection Signature) identified on chromosome 25 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome Selection signature (SS) identified in the red rectangle without genes annotations. Outside of rectangle at distances more than 1 Mb no genes are shown.

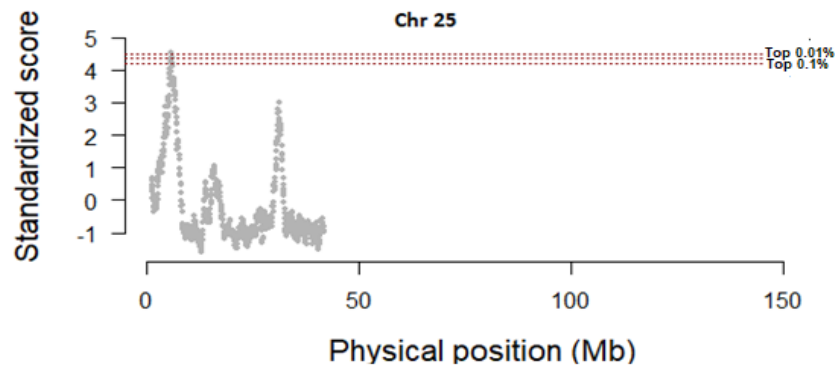


Figure 9(a). Physical position in megabases (Mb) on chromosome 25 of the genomic region according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

On BTA29, one SS between 44,707,685 and 44,796,094 bp, was identified (Table 2, Figure 10, Figure 10 (a)) with annotations for the *SPTBN2* (Spectrin beta chain) and *RBM4* (RNA-binding protein 4) genes. (Ensembl, 2019) (Table 2, Annex 1).

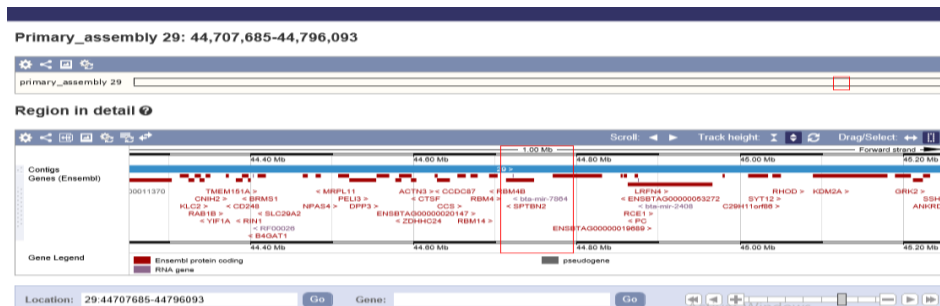


Figure 10. Genomic region (Selection Signature) identified on chromosome 29 of Blanco Orejinegro (BON) and Sanmartinero (SM) genome Selection signature (SS) identified in the red rectangle has annotations for the *SPTBN2* and *RBM4* genes. Indicating that these SS and SNPs within it, are highly associated to these gene.

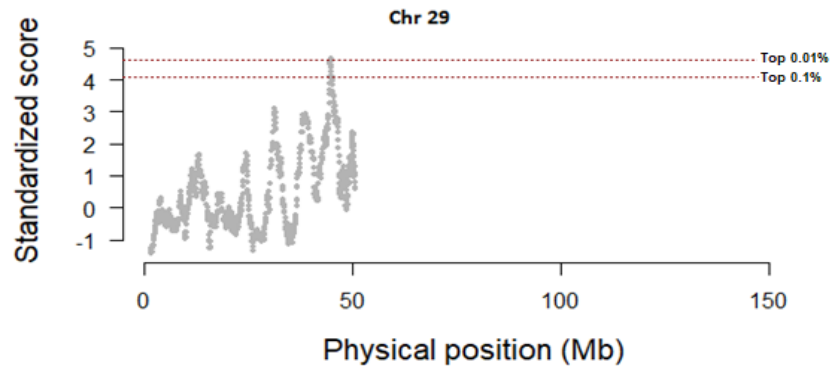


Figure 10(a). Physical position in megabases (Mb) on chromosome 29 of the genomic region according to the value of the standardized linkage disequilibrium on percentile 0.1 and 0.01. Figure made in R program.

DISCUSSION

Study of LD variation patterns (varLD) throughout genome, allows to identify candidate regions (signatures of selection) under positive selection and genes related to productive and adaptation performance of Colombian creole cattle breeds. varLD method has great potential to capture recent selection that manifests itself in high LD values in narrow regions throughout the genome undergo selective sweep (Pérez O'Brien et al., 2014). Studies on LD in livestock populations have shown that LD is much more extensive than humans. Human populations analyzed (Ardile et al., 2002; Hee, 2012; Wall and Pritchard, 2003), had shown very similar extensions and patterns of LD, with very few differences only in limited regions. Cattle populations differ from human ones because they have an effective smaller population size and experiencing strong recent selection due to breed formation and because are constantly subjected to artificial selection and advanced reproductive technologies in order to overcome the productive performance of milk and meat, adaptation and reproductive performance (Khatkar et al., 2008; McKay et al., 2007), which rise LD values. This makes LD comparisons between cattle breeds worthwhile (Bovine Hapmap consortium, 2009). In the absence of selection, LD breaks down rapidly to values below 0.3, while natural or artificial selection of genome regions and beneficial alleles, increases LD between hitchhiking alleles (Li et al., 2017). In general, LD decay pattern is that it decays at distances of 100 Kb. Bejarano, (2016) in his study of LD decay in the Creole breeds ROMO and BON showed LD average of 0.3 ($r^2 = 0.3$) in the ROMO breed at a distance of 100Kb, while the BON breed showed a faster fall, reaching this same value at a distance of 70Kb. In other *Bos taurus* breeds, to milk and meat production, have been reported $r \geq 0.3$ at distances less than or equal to 30 Kb (Bolormaa et al., 2011b; Larmer et al., 2014), however, in Holstein Sargolzaei et al. (2008) found optimal LD values at distances close to 100 Kb, which coincides with the results found in the Bejarano study (2016) for BON and ROMO Creole breeds.

In the current study, high LD values in some regions were identified (signatures of selection) on chromosomes 3, 5, 11, 15, 18, 23, 25 and 29 with standardized varLD values between 4.01 and 5.24 in percentiles 0.1 and 0.01. only on BTA3, 5,18, 21 and 29, genes of exact location within these regions were identified, while in the rest of the chromosomes

genes outside of these regions, were located, some of them very near of the SS (BTA23), others very far (BTA21) and in one region, no annotations for genes (BTA25).

On BTA3, within the SS identified, annotations for the *SYT6* gene (Synaptotagmin 6) were found. This gene acts in the synapse and dopamine regulation (Uniprot, 2019). Dopamine is a precursor of adrenaline and norepinephrine involved in the pathophysiology of stress in the activation of the adrenal hypothalamic axis (Guyton and Hall, 2011). This gene has been reported for its activity in exocytosis of human sperm acrosome (Michaut et al., 2001) and human pulmonary synovial sarcoma (Li et al., 2019), but there are no reports in cattle, which may be is involved in both, reproductive processes and responses to stressful environmental situations. Surrounding this SS, around 10 Kb, *OLFML3* and *HIPK1* genes were identified.

OLFML3 (Olfactomedin-like protein 3) is involved in embryo and placental development, indicating its participation in reproductive processes. This gene has been reported in prenatal skeletal muscle development in pigs (Zhao et al., 2012), in glaucoma in dogs Border Collie breed (Pugh et al., 2019) and embryonic development (Zeng et al., 2005). *HIPK1* gene (Homeodomain interacting protein kinase 1) has been reported acting in cellular productivity may be in protein production (Inwood et al., 2018). These two genes are possibly associated to this SS. *OLFML3*, for its activity in early embryonic development and skeletal muscle, is possibly involved in productive and reproductive mechanisms in cattle. *HIPK1*, is possibly involved in mechanisms of production of meat and milk in cattle. In another study on signatures of selection in Red Polish cattle breed using the paired distance methodology *Fst*, Gurgul et al. (2019) identified one SS on BTA3 at 32.5 to 32.9 Mb very close to the identified region in the current study and within a large region in strong linkage disequilibrium with extended haplotypic structure indicating reduced genetic variation caused by pressure of selection. De Leon et al. (2019), identified two regions associated to adaptability traits on chromosome 3 in BON and SM breeds, the region 66,486,992-66,554,349 bp in BON breed and the region 83,833,592-83,952,776 bp in SM breed, both regions associated to the coefficient of adaptability (CA). In these regions they found annotations for the *GIPC2* and *TM2D1* genes respectively. *TM2D1* gene is involved in the physiological response to heat stress.

On BTA5, one SS between 55,085,241 and 55,582,258 bp was identified. Within this region annotations for the *ATP23* gene were found and near of the SS annotations for *CTDSP2* gene (CTD (Carboxy-terminal domain small phosphatase 2) were found. *ATP23* gene has metalloendopeptidase activity, it has been reported in yeasts, humans and other mammals in protease activity, is involved in proteolysis of intracellular proteins as well as in the respiratory chain of the mitochondria fulfilling multiple cellular functions in which temperature sensitive alleles are involved to affect the conversion of oxygen into water by enzyme cytochrome oxidase causing deficiency in the respiratory chain. This deficiency is corrected by mutations of *COX* gene that codes for cytochrome oxidase; mechanism associated to *ATP23* gene (Zeng et al., 2007). By this fact this gene is possibly involved in adaptive processes. Pitt et al. (2019) identified one signature of selection on BTA5 between 195.000.000 and 200.000.000 bp in their study on signatures of selection in the Colombian creole cattle breeds Costeño con Cuernos, Romosinuano y Sanmartinero, using the *Fst* paired distance methodology in which annotations for the *ATP2B1* gene they found, according to them, this gene is involved in adaptive mechanisms of tick resistance. *CTDSP2* gene of phosphatase activity is crucial for cell cycle progression. Kloet et al. (2015)

reported this gene in mice and humans whose expression is controlled by FOXO cells (Forkhead box), involved in inhibition of cancer in humans and cellular energy homeostasis, possibly involved in productive mechanisms. Pérez O'Brien et al. (2014), reported on BTA5, but in a different region at 48.5-49.1 Mb using varLD methodology a selective sweep in the Gyr breed, a specific repressor of transforming growth factor beta receptor (TGF-beta) involved in negative regulation of the cell differentiation of the skeletal muscle. Igoshin et al. (2019) identified signatures of selection and genes associated with maintenance of body temperature in Siberian cattle during cold stress using the combination methodology of four statistics (Haplotypic homozygosity H1, Haplotypic homozygosity H12, Tajima's D index and Pi nucleotide diversity) taking the p-value from correlation of the correlation matrix between the 4 statistics and converted into q-values. They identified one SS on BTA5 and *MEF2A* and *NBEA* genes. *MEF2A* (myocyte enhancer factor 2A) is involved in acclimatization in fish and *NBEA* (Neurobeachin) associated with maintaining of body temperature between normal ranges during heat stress.

In others varLD studies in Italian bovine cattle *Bos taurus* Marchigiana and Piemontese, Sorbolini et al. (2015) reported SS with varLD standardized ≥ 4 on BTA5, BTA11, BTA15, BTA18 and BTA23. In another study of varLD in *Bos indicus* and *Bos taurus* beef and milk cattle, Pérez O'Brien et al. (2014) showed signatures of selection on chromosomes BTA3 and BTA5 with varLD standardized between 4.5 and 5.4 in the 0.1th percentile and between 5.6 and 7.6 in the 0.01th percentile. Kukučková et al. (2017) determined candidate regions under positive selection in European bovine *Bos taurus*, Holstein, Simmental, Charolais among others using varLD methodology, with varLD standardized ≥ 4 in 0.1 and 0.01 percentiles on chromosomes 7, 9, 11, 14, 20 and 24. In other different species like swines, Li et al. (2017) determined SS by varLD methodology on chromosomes 7, 9, 13 and 14 in the percentile 0.1. Urbinati et al. (2016) determined SS in beef cattle Brazilian Canchim breed by iHS method, showing conserved genomic regions on chromosomes 5 and 14. These results can explain the conserved trend of chromosomes specific to show positive selection more than the rest.

On BTA11 no annotated genes in the SS identified were found. However, at 200 Kb, the *LRR1MI* gene (Leucine-rich repeat transmembrane neuronal protein 1) was identified between 55.211.496 and 55.214.132 bp possibly associated to this SS. This gene has been reported in humans by Beste et al. (2018), expressed almost exclusively in brain involved in synapse organization, at pre- and post-synapse level, in the control of cognitive processes, functional brain asymmetry and schizophrenia. According to Oczkowicz et al. (2018), this gene is of biallelic expression in brain of pigs. Voikar et al. (2013) reported this gene involved in neuropsychiatric syndrome in mice, expressing in thalamus, hippocampus and limbic cortex, areas of the brain that generate abnormal behaviors in response to stressful environmental changes such as claustrophobia. Possibly this gene is involved in adaptive mechanisms in cattle. Pitt et al. (2019) identified one signature of selection on BTA11 between 105.500.001 and 106.000.000 bp using the *Fst* paired distance methodology, in which annotations for the *LRR26* and *SLC34A3* genes involved in adaptive mechanisms. Kukučková et al (2017) determined signatures of selection by the varLD methodology in different European cattle breeds reporting one SS on BTA11 at 59.04-63.13 Mb close to the SS identified in the current study, identifying the *MAP3K7* gene at the 0.01th percentile, associated to feed efficiency and growth, which indicates their participation in productive mechanisms.

On BTA 15 one SS at 24.184.623-24.186.798 bp was identified, no annotations for genes were found. However, near this region at 100 Kb, the *ZW10* gene (Centromere / kinetochore protein zw10) was identified among 24.284.251 and 24.319.569 bp, possibly linked to this SS. This gene has been associated to congenital heart disease (Sun et al., 2017) and Roberts syndrome in humans (Musio et al., 2004) of anchoring chromosomes activity, chromosomal movements during cell division, movement and segregation of chromosomes (Uniprot, 2019). There no reports of this gene in cattle.

On BTA 18, one SS was identified. Within this region annotations for the *SLC6A2* (Transporter) and *CESI* (Carboxylic ester hydrolase) were found. *SLC6A2* transports dopamine inside neurons, it is highly expressed in lung. Dopamine is a neurotransmitter catecholamine and metabolic precursor of noradrenaline and adrenaline, hormones present in response to stress (Uniprot, 2019). Many genes expressed in lung are related to response to stressful situations. Indeed, Zhang et al. (2016) reported polymorphisms of *SLC6A2* gene associated to post-traumatic stress in human veteran war fighters. Polymorphisms and methylation of this gene have been associated to suicide in human patients, panic and deep depressive syndrome (Bayles et al., 2013; Kim et al., 2014). Dikmen et al. (2013) in their study of genomic association in American Holstein cows undergo heat stress, reported *SLCO1C1* gene (solute carrier, organic anion transporter, member 1C1) on BTA5, gene of solute transporter family, in humans, *SLCO1C1* mediates thyroxine and triiodothyronine transport. This gene possibly acts in response to environmental stress in cattles. The second gene identified in this SS was *CESI*, involved in the metabolism of lipids. Sorbolini et al. (2015) reported the *LCAT* gene (Lecithin-cholesterol acyltransferase) identified by varLD method on chromosome 18, in Italian beef cows *Bos taurus*, involved in cholesterol homeostasis.

Maiorano et al. (2018), determined signatures of selection by the iHS, Fst and XP-EHH methodologies in Gyr breed in Brazil in the two subtypes of this breed of milk and meat production. In a significant XP-EHH genomic region on BTA18 at 54.862.897-54.863.397 bp pretty far from the SS identified in the current study, they identified the *SLC8A2* (solute transporter family) gene similar to the *SLC6A2* gene identified in the current study, involved in responses to stressful environmental situations, therefore in adaptive mechanisms.

On BTA21, three SS were identified, with annotations for the *CFAP161* (Cilia and flagella associated protein 161), *SSTR1* (Somatostatin receptor 1) and *CLEC14A* (C-type lectin domain family 14, member A) genes. *CFAP161* was identified near the first SS. There are few reports of this gene in the literature, McClintock et al. (2008) showed this gene involved in sperm motility and fluid movements of mice and other mammals including human being, possibly involved in reproductive mechanisms in cattles. In the second SS, within it, the *SSTR1* gene of somatostatin receptor activity was identified, a peptide hormone that regulates endocrine system. This gene was reported by Zhao et al. (2018) which codes for the somatostatin receptor involved in productive growth characteristics in New Zealand's Romney sheep, because it plays a significant role in growth hormone regulation. Productive traits of weight at birth, weaning weight and carcasses weight were assessed by these researchers, they showed affecting these traits, favorably and unfavorably according to the allelic variants of this gene. In another study, Wang et al. (2006), reported ablation in this gene in mice resulting in alteration of glucose homeostasis due to an exaggerated increase in insulin levels. By these facts this gene is a strong candidate gene for

productive mechanisms in cattle. Finally, close to the third SS identified the *CLECI4A* gene (C-type lectin domain family 14, member A) was found. This gene is involved in the formation of new blood vessels (angiogenesis) (Uniprot, 2019). Li et al. (2017) identified aberrant methylation of DNA in epigenetic signatures of selection in human patients with uterine carcinosarcoma. The hypermethylation of *CLECI4A* gene promoters is associated to tumors vascularization. There are no reports about function of *CLECI4A* gene in cattle, possibly involved in productive mechanisms.

Sevane et al. (2018) in their study about the role of epigenetic regulation on the adaptive processes in Colombian Creole cattle breeds and their Iberian ancestors comparing the methylomas between these breeds, reported 12 genes in a hypermethylated region on BTA21 of Costeño con cuernos and Sanmartinero breeds, related to cellular membrane components. Many of these hypermethylated genes are involved in immune processes against infections and parasites in adaptive mechanisms. Pitt et al. (2019) identified two signatures of selection on BTA21, in the Colombian creole cattle breeds Costeño con Cuernos, Romosinuano and Sanmartinero, using the paired distance methodology, *Fst*. The first one between 16,500,000 and 17,000,000 bp and the second one between 29,500,000 and 30,000,000 bp. In the first SS, annotations for the *U6* and *KLHL25* genes they found. These genes are involved in productive, reproductive and tuberculosis susceptibility processes. In the second SS, they identified *PCSK6* and *SNRPA1* genes, involved in reproductive processes.

On BTA23, one SS between 39.339.694 and 39.354.709 bp was identified, no annotations for genes. However, very close to this region at 39.355,962-39.387.656 bp the *DEK* (DEK protein) and *KDM1B* genes (Lysine demethylase 1B), at 39.389.854-39.432.004 bp were identified. *DEK* gene acts like DNA repair. *KDM1B* gene of zinc finger activity, a protein that allows DNA-amino acids binding to then initiate transcription by RNA polymerase. This gene also has oxidoreductase activity in response to oxidative stress (Uniprot, 2019). These two genes are possibly associated with this SS. *DEK* gene is a widely reported oncogene in human liver carcinomas (Lee et al., 2019), suppression of apoptosis and development of lung tumors (Zu et al., 2017), promoter of cell metabolism and glycolysis (Matrka et al., 2017) and as a regulator of hematopoiesis (Capitano and Broxmeyer, 2017). There is no evidence of this gene in cattle. But by involved in DNA repair, possibly acts in response to oxidative stress. The *KDM1B* gene is widely reported in humans and animals. In humans it has been associated with breast cancer (Chen et al. 2017). Fellous et al. (2019) showed this gene involved in methylation of the histones, development and reproduction of mangrove fishes that self-fertilize. (Cui et al., 2018) showed the association of this gene with testis weight in piglets. As noted here, the *KDM1B* gene is involved in different carcinogenic processes in humans, and response to oxidative stress and reproductive mechanisms in animals, which makes it a candidate gene to play an important role in adaptation and reproduction processes.

On BTA29 one SS was identified and two genes contained in it. *SPTBN2* gene (Spectrin beta chain) and *RBM4* gene (RNA-binding protein 4). *SPTBN2* was reported by Wang et al. (2014) involved in cerebellar ataxia in humans. Forman et al. (2012) identified a mutation (deletion) of this gene involved in cerebellar neonatal degeneration in dogs, Beagle breed. In bovines there are no reports on this gene, possibly negatively affects muscle development. The *RBM4* gene is reported by Lin et al. (2013) of endocrine activity in promoting the differentiation of pancreatic cells (islets of Langerhans) and differentiation

of muscle cells. Although there are no reports in cattle, both activities in this gene possibly affect productive performance. De Leon et al. (2019), identified several regions associated to reproductive traits on chromosomes 5, 11, 15, 18, 23, 25, 29 in BON and SM breeds. In these regions they found annotations for the *SLC6A16*, *TMEM50A*, *NLRP9*, *KHDRBS2* and *UBE2C* genes associated to reproductive performance.

On the other hand, Martínez et al. (2013) studied the adaptability processes of the BON breed in a study of genetic variability using microsatellites using the principal component methodology by measuring Nei's standard genetic distances between BON breed populations. They found lower values of genetic distance between populations that were geographically close and had recently shared bulls. Their results indicated that these populations shared many alleles and that the frequency of alleles shared between them were similar, suggesting lower genetically distant populations.

CONCLUSIONS

The high var LD score between genomic regions obtained in this work indicate strong recent selection acting in these regions for adaptive, productive and reproductive purposes because selected beneficial alleles are fixed in the population increasing their frequency, as demonstrated in the 10 regions resulting from the standardized VarLD on chromosomes 3, 5, 11, 15, 18, 21, 22, 23, 25 and 29. In most of these regions, annotations were found for genes involved in biological processes. On region 29,224,886 bp on chromosome 3, annotations for the *OLFM3* gene were found involved in reproductive traits. In another region, 23,903,882 bp on chromosome 18, annotations for the *SLC6A2* gene were found, involved in the response to high stress. Also, we identified the genes *CTDSP2*, *CES1*, *CFAP161*, *CLEC14A*, *HIPK1*, *RBM4*, *SSTR* involved in the expression of economic traits.

In processes where the natural or artificial selection does not act, LD breaks quickly through the generations decreasing the score. This study helps to explain the genetic architecture conformed for productive, reproductive and adaptation processes. The different genes identified in these selected regions consistent with other research work, are involved in the various biological processes from production, reproduction and adaptation and can be used in selection and conservation programs of these Colombian Creole breeds.

ACKNOWLEDGMENTS

This Research was supported by the Ministerio de Agricultura y Desarrollo Rural under Agreement 1936 TV19 of 2019. We thank all the stockbreeders of BON and SM who provided the samples necessary to carry out this study.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Alachiotis N and Pavlidis P (2016). Scalable linkage-disequilibrium based selective sweep detection: a performance guide. *Gigascience* 5: 7.

- Agrosavia Corporación Colombiana de Investigación Agropecuaria. <http://www.agroavia.co/agrosavia>. Accessed 17 October 2020.
- Ardlie KG, Kruglyak L and Seielstad M (2002). Patterns of linkage disequilibrium in the human genome. *Nature Rev. Genet.* 3: 299-309.
- Bayles R, Baker EK, Jowett JBM, Barton D, et al. (2013). Methylation of the SLC6a2 gene promoter in major depression and panic disorder. *PLoS One.* 8: e83223.
- Bejarano GDH (2016). Genomic association study for growth traits in Creole cattle breeds Blanco Orejinegro and Romosinuano. Master of Science Thesis, *Universidad Nacional de Colombia, Bogotá*.
- Beste CH, Arming LN, Qding WM, Epplen JT, et al. (2018). Cognitive Control Processes and Functional Cerebral Asymmetries: Association with Variation in the Handedness-Associated Gene *LRRMT1*. *Mol. Neur.* 55: 2268-2274.
- Biswas S and Akey JM (2006). Genomic insights into positive selection. *Trends Genet.* 22: 437-46.
- Bolormaa S, Hayes BJ, Savin K, Hawken, R et al. (2011). Genome-wide association studies for feedlot and growth traits in cattle. *J. Animal Sci.* 89: 684-1697. doi: 10.2527/jas.2010-3079.
- Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science.* 324: 528-532.
- Capitano ML and Broxmeyer HE (2017). Review. A role for intracellular and extracellular *DEK* in regulating hematopoiesis. *Curr. Opin Hematol.* 24: 300-306. doi: 10.1097/MOH.0000000000000344.
- Chen L, Vasilatos SN, Qin, Y Katz TA, et al. (2017) Functional characterization of lysine-specific demethylase 2 (*LSD2/KDM1B*) in breast cancer progression. *Oncotarget.* 8: 81737-81753. doi: 10.18632/oncotarget.19387.
- Cui Y, Hu T, Chen R, Yu SDW, et al. (2018). Novel 17-bp Deletion in *KDM1B* Gene is Significantly Associated with Testis Weight in Male Piglet. *Anim. Biotechnol.* 29: 252-258. doi: 10.1080/10495398.2017.1370427.
- De León C, Manrique C, Martínez R, Rocha JF (2019). Genomic association study for adaptability traits in four Colombian cattle breeds. *Genet. Mol. Res.* 18(3): GMR18373. <https://doi.org/10.4238/gmr18373>.
- Dikmen S, Cole JB, Null DJ and Hansen PJ (2013). Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. *PLoS One.* 8: e69202. doi: 10.1371/journal.pone.0069202.
- Ensembl online genome data base BioMart Tool. <http://www.ensembl.org/biomart/martview/>. Accessed 17 October 2020.
- Ensembl Genome Browser. <http://www.ensembl.org/index.html>. Accessed 17 October 2020.
- Fellous A, Earley RL and Silvestre F (2019). The Kdm/Kmt gene families in the self-fertilizing mangrove rivulus fish, *Kryptolebias marmoratus*, suggest involvement of histone methylation machinery in development and reproduction. *Gene.* 687: 173-187. doi: 10.1016/j.gene.2018.11.046.
- Forman OP, De Risio L, Stewart J, Mellersh CS, et al. (2012). Genome-wide mRNA sequencing of a single canine cerebellar cortical degeneration case leads to the identification of a disease associated *SPTBN2* mutation. *BMC genetics.* 13: 55
- Gene Cards Human Gene Database. In: Weizmann Inst. Sci. <http://www.genecards.org/>. Accessed 17 October 2020.
- Gurgul A, Jasielczuk I, Semik-Gurgul E, Szmatola T, et al. (2019). Diversifying selection signatures among divergently selected subpopulations of Polish Red cattle. *J. Appl. Genet.* 60: 87-95. <https://doi.org/10.1007/s13353-019-00484-0>.
- Guyton A and Hall JE (2011). Tratado de Fisiología médica. 12th Edición. Elsevier. p. 987.
- Hee OT (2012). Population diversity as quantified by inter-population variation in patterns of linkage disequilibrium. (Doctoral dissertation). *Singapore University. Saw Swee Hock School of Public Health*. Illumina Inc. BovineSNP50 Genotyping BeadChip. In: http://www.illumina.com/documents/products/datasheets/datasheet_bovine_snp50.pdf. Accessed 17 October 2020.
- Inwood S, Buehler E, Betenbaugh M, Lal M, et al. (2018). Identifying *HIPK1* as Target of miR-22-3p Enhancing Recombinant Protein Production from HEK 293 Cell by Using Microarray and HTP siRNA Screen. *Biotechnol. J.* 13: 10.1002/biot.201700342.
- Khatkar MS, Nicholas FW, Collins AR, Zenger KR, et al. (2008). Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics.* 9: 187. doi: 10.1186/1471-2164-9-187.
- Kim Yong-Ku, Hwang Jung-A, Lee Heon-Jeong, Yoon Ho-Kyoung, et al. (2014). Association between norepinephrine transporter gene (SLC6A2) polymorphisms and suicide in patients with major depressive disorder. *J. Affective Disor.* 158: 127-132.
- Kloet DEA, Polderman PE, Eijkelenboom A, Smits LM, et al. (2015). FOXO target gene *CTDSP2* regulates cell cycle progression through Ras and p21(Cip1/Waf1). *J. Biochem. J.* 469: 289-298.
- Krzanowski WJ (1993). Permutational test for correlation matrices. *Stat. Comput.* 3: 37-44.
- Kukučková V, Moravčíková N and Kasarda R (2017). Variation in Linkage Disequilibrium Patterns between Populations of Different Production Types. *Agric. Conspec. Sci.* 82:105-109.
- Larmer SG, Sargolzaei M and Schenkel FS (2014). Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. *J. Dairy Sci.* 97: 3128-3141. doi: 10.3168/jds.2013-6826.

- Lee SY, Jung W, Lee JKA, Kim HK, et al. (2019). High expression of *DEK* gene is associated with poor prognosis in hepatocellular carcinoma. *Histol. Histopathol.* 34: 1279-1288. doi: 10.14670/HH-18-125.
- Li X, Yang S, Dong K, Tang Z, et al. (2017). Identification of positive selection signatures in pigs by comparing linkage disequilibrium variances. *Anim. Genet.* 48: 600-605. doi: 10.1111/age.12574.
- Li J, Xing X, Li D, Zhang B, et al. (2017). Whole-Genome DNA Methylation Profiling Identifies Epigenetic Signatures of Uterine Carcinosarcoma. *Neoplasia.* 19: 100-111. doi: 10.1016/j.neo.2016.12.009.
- Li X, Wu D, Zheng Y, Yang H, et al. (2019). Poorly differentiated pulmonary synovial sarcoma with *SYT* gene amplification: A case report. *Mol. Clin. Oncol.* 10: 249-252.
- Lin JC, Yan YT, Hsieh WK, Peng J, et al. (2013). *RBM4* promotes pancreas cell differentiation and insulin expression. *Mol. Cell. Biol.* 33: 319-27. doi: 10.1128/MCB.01266-12.
- Maiorano AM, Lourenco DL, Tsuruta S, Toro AM, et al. (2018). Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. *PLoS One.* 13: e0200694. doi.org/10.1371/journal.pone.0200694.
- Martinez SRA, Gallego GJL, Gómez VY, Moreno LF, et al. (2012). Eficiencia productiva de la raza BON en el trópico colombiano. *Editorial Salmon D.C.*
- Matrka MC, Watanabe M, Muraleedharan R, Lambert PF, et al. (2017). Overexpression of the human *DEK* oncogene reprograms cellular metabolism and promotes glycolysis. *PLoS One.* 12: e0177952. doi: 10.1371/journal.pone.0177952.
- McClintock TS, Glasser CE, Bose SC and Bergman DA (2008). Tissue expression patterns identify mouse cilia genes. *Physiological genomics.* 32: 198-206.
- McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, et al. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genet.* 8: 74. doi: 10.1186/1471-2156-8-74.
- Michaut M, De Blas G, Tomes CN, Yunes R, et al. (2001). Synaptotagmin VI participates in the acrosome reaction of human spermatozoa. *Dev. Biol.* 235: 521-529. doi:10.1006/dbio.2001.0316.
- Montgomery S (2008). Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Rev. Genet.* 9: 477-485
- Msalya G, Kim ES, Laisser LKE, Kipanyula MJ, et al. (2017). Determination of Genetic Structure and Signatures of Selection in Three Strains of Tanzania Shorthorn Zebu, Boran and Friesian Cattle by Genome-Wide SNP Analyses. *PLoS One.* 12: e0171088. doi: 10.1371/journal.pone.0171088.
- Musio A, Mariani T, Montagna C, Zamboni D, et al. (2004). Recapitulation of the Roberts syndrome cellular phenotype by inhibition of *INCENP*, *ZWINT-1* and *ZW10* genes. *Gene.* 331: 33-40.
- Oczkiewicz M, Zszmatola T, Piorkowska K and Ropka-Molik K (2018). Variant calling from RNA-seq data of the brain transcriptome of pigs and its application for allele-specific expression and imprinting analysis. *Gene.* 641: 367-375 <https://doi.org/10.1016/j.gene.2017.10.076>.
- Perez O'Brien, Utsunomiya YT, Meszaros G, Bickhart DM, et al. (2014). Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet. Sel. Evol.* 46: 19. doi: 10.1186/1297-9686-46-19 PMID: 24592996.
- Pitt D, Bruford MW, Barbato M, Orozco-terWenge P, et al. (2019). Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evol Appl.* 12: 105-122.
- Pugh CA, Farrell LL, Carlisle AJ, Bush SJ, et al. (2019). Arginine to Glutamine Variant in Olfactomedin Like 3 (*OLFML3*) Is a Candidate for Severe Goniodysgenesis and Glaucoma in the Border Collie Dog Breed. *G3 (Bethesda, Md.)* 9: 943-954.
- Purcell S, Neale B, Todd-Brown K, Thomas L, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81: 559-575. doi: 10.1086/519795.
- Sabeti PC, Pardis C, Sabeti DE, Reich JM, et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature.* 419: 832.
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, et al. (2006). Positive natural selection in the human lineage. *Science.* 312: 1614-1620.
- Sabeti PC, Varilla P, Fry B, Lohmueller J, et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature.* 449: 913.
- Sargolzaei M, Schenkel FS, Jansen GB and Schaeffer LR (2008) Extent of linkage disequilibrium in Holstein cattle in North America. *J. Dairy Sci.* 91: 2106-2117. doi: 10.3168/jds.2007-0553.
- Sargolzaei M, Chesnais JP and Schenkel FS (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics.* 15: 478. doi: 10.1186/1471-2164-15-478.
- Sevane N, Martinez R and Bruford MW (2018). Genome-wide differential DNA methylation in tropically adapted Creole cattle and their Iberian ancestors. *Anim. Genet.* 50: 15-26. doi: 10.1111/age.12731.
- Sorbolini S, Marras G, Gaspa G, Dimauro C, et al. (2015). Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories. *Genet. Sel. Evol.* 47: 52. doi 10.1186/s12711-015-0128-2.
- Sun Chao-Yu, Sun Chi, Cheng Rui, Shi Shuai, et al. (2017). Rs2459976 in *ZW10* gene associated with congenital heart diseases in Chinese Han population. *Oncotarget.* 9: 3867.

- Teo YY, Fry AE, Bhattacharya K, Small KS, et al. (2009). Genome-wide comparisons of variations in linkage disequilibrium. *Genome Res.* 19: 1849-1860.
- Uniprot (2019). UniProt. <http://www.uniprot.org/>. Accessed 17 January 2021.
- Urbinati I, Stafuzza NB, Oliveira MT, Seleguim TC, et al. (2016). Selection signatures in Canchim beef cattle. *J. Animal Sci. Biotechnol.* 7: 29.
- Voight BF, Kudaravalli S, Wen X and Pritchard JK (2006). A map of recent positive selection in the human genome. *PLoS Biology* 4: e72. doi: 10.1371/journal.pbio.0040072.
- Voikar V, Kuleskaya N, Laakso T, Lauren J, et al. (2013). *LRRTM1*-deficient mice show a rare phenotype of avoiding small enclosures--a tentative mouse model for claustrophobia-like behavior. *Behav. Brain Res.* 238: 69-78. doi: 10.1016/j.bbr.2012.10.013.
- Wall JD and Pritchard JK (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nature Review Genetics.* 4: 587.
- Walsh B and Lynch M (2018). Evolution and Selection of Quantitative Traits. *Oxford University Press*.
- Wang Y, Koh K, Miwa M, Yamashiro N, et al. (2014). A Japanese SCA5 family with a novel three-nucleotide in-frame deletion mutation in the *SPTBN2* gene: a clinical and genetic study. *Am. J. Human Genet.* 59: 569.
- Wang XP, Norman M, Yang J, Magnusson J, et al. (2006). Alterations in glucose homeostasis in *SSTR1* gene-ablated mice. *Mol. Cell. Endocrin.* 247: 82-90.
- Zaykin DV, Zhaoling MZ and Ehm MG (2006). Contrasting Linkage-Disequilibrium Patterns between Cases and Controls as a Novel Association-Mapping Method. *Am. J. Hum. Genet.* 78: 737-746
- Zeng LC, Han ZG and Ma WJ (2005). Elucidation of subfamily segregation and intramolecular coevolution of the olfactomedin-like proteins by comprehensive phylogenetic analysis and gene expression pattern assessment. *FEBS Lett.* 579: 5443-5453.
- Zeng X, Neupert W and Tzagoloff A (2007). The Metalloprotease Encoded by *ATP23* Has a Dual Function in Processing and Assembly of Subunit 6 of Mitochondrial ATPase. *Mol. Biol. Cell.* 18: 617-626.
- Zhang J, Mandel H, Hamner M and Wang Z (2016). Norepinephrine transporter gene *SLC6A2* polymorphisms and promoter methylation in peripheral blood of veterans with posttraumatic stress disorder. *Biol. Psych.* 59: 51S-51S.
- Zhao JZ, Xinhua HLZ, Wuang N, Zhonglin T, et al. (2012). *OLFML3* expression is decreased during prenatal muscle development and regulated by microRNA-155 in pigs. *Int. J. Biol. Sci.* 8: 459.
- Zhao F, Zhou H, Li S, Fang Q, et al. (2018). Growth and carcass trait association with variation in the somatostatin receptor 1 *SSTR1* gene in New Zealand Romney sheep. *J. Agri. Res.* 61: 477-486.