Frontiers in Bioscience Scholar

Front. Biosci. (Schol Ed) **2024**; 16(3): 18
https://doi.org/10.31083/j.fbs1603018

Original Research

SNP Markers: Analysis of Genetic Diversity and Identification of Genomic Regions in Pantaneiro Sheep and Texel Sheep Under Natural Selection

Layonel Alves de Sousa[1,*], Adriana Mello de Araujo[2], Francisco de Assis Diniz Sobrinho[3], Miklos Maximiliano Bajay[4], Raquel Soares Juliano[2], Fernando Alvarenga Reis[5], Marcos Barbosa Ferreira[6], Gustavo Daniel Vega-Britez[7], Fernando Miranda de Vargas Junior[7]

[1]Department of Veterinary Medicine, Center for Agroveterinary Sciences, State University of Santa Catarina, 88520-000 Lages, Santa Catarina, Brazil
[2]Embrapa Pantanal, Brazilian Agricultural Research Company, 79320-900 Corumbá, Mato Grosso do Sul, Brazil
[3]Department of Teacher Education, Federal University of Piaui, 64049-550 Teresina, Piauí, Brazil
[4]Department of Fisheries Engineering and Biological Sciences, Center for Higher Education in the Southern Region, State University of Santa Catarina, 88790-000 Laguna, Santa Catarina, Brazil
[5]Embrapa Caprinos e Ovinos, Brazilian Agricultural Research Company, Núcleo Regional Centro-Oeste, 70770-901 Campo Grande, Mato Grosso do Sul, Brazil
[6]Postgraduate Program in Sustainable Agribusiness, Anhanguera-Uniderp University, 79003-010 Campo Grande, Mato Grosso do Sul, Brazil
[7]Faculty of Agricultural Sciences, Federal University of Grande Dourados, 79825-070 Dourados, Mato Grosso do Sul, Brazil
*Correspondence: Layonel.sousa@edu.udesc.br (Layonel Alves de Sousa)
Academic Editor: Gustavo Caetano-Anollés
Submitted: 10 June 2024 Revised: 22 August 2024 Accepted: 12 September 2024 Published: 29 September 2024

Abstract

Background: Sheep farming is growing substantially in Brazil, driven by the increasing demand for sheep meat. This rising demand has heightened the focus on sheep, making them the subject of numerous studies, including those centered on genetic analysis. A notable research area involves Pantaneiro sheep, which are indigenous to the Pantanal region of Mato Grosso do Sul and other locations. These sheep are of particular interest due to their adaptation to the unique environmental conditions of the Pantanal, a floodplain characterized by its distinctive climatic and ecological features. This study primarily aimed to conduct a comprehensive genomic analysis of Pantanal sheep subjected to natural selection within the Pantanal region and compare different sample herds using methodological approaches. **Methods**: Genomic analysis was performed to examine genetic diversity and structure via GGP50K single nucleotide polymorphism (SNP) analysis. A sample of 192 adult sheep over 4 years old was categorized into seven populations based on location: Six populations comprised Pantaneiro sheep with one Texel sheep population. Outlier SNPs were assessed to pinpoint regions under natural selection, with comparisons between the Pantaneiro and the commercial Texel breeds. All data analyses were conducted using the R programming language, employing specialized genetic analysis packages. These outlier SNPs were detected using three methodologies, PCAdapt, OutFLANK, and FDIST2/fsthet, with false discovery rate (FDR) corrections applied to ensure result accuracy. Each method was evaluated, and the genes associated with the identified SNPs were cross-referenced with the most recent sheep genome database, focusing specifically on genes with known phenotypic traits. **Results**: Analysis of a sample comprising 192 adult individuals revealed greater genetic variability within the Pantaneiro breed than the Texel breed, highlighting the adaptation of the Pantaneiro breed to the unique Pantanal environment. Conversely, the Texel breed exhibited significantly higher levels of inbreeding, attributed to its controlled breeding practices. Outlier SNPs were detected with notable variation across different methodologies, underscoring the importance of FDR correction in ensuring the reliability and concentration of identified outliers. These outlier SNPs facilitated the identification of genes associated with key phenotypic traits, including hair growth, tissue regeneration, pigmentation regulation, and muscle capacity. **Conclusion**: The integrated analysis of methodologies demonstrated significant efficiency in elucidating the genomic landscape of Pantanal sheep, highlighting the genetic richness inherent in sheep from the Pantanal region of Mato Grosso do Sul. The techniques employed effectively identified outlier SNPs associated with phenotypically relevant genes. These findings, which reveal greater genetic variability and adaptability, underscore the potential of these animals for future research and their significance within Brazilian sheep farming. The Texel breed served as a valuable comparative group, illustrating the limited genetic variability in highly controlled breeding environments.

Keywords: sustainability; selection signatures; conservation

1. Introduction

Pantaneiro sheep are ruminant animals found in the Pantanal biome, predominantly in the southern region of the Brazilian Pantanal in Mato Grosso do Sul. Pantaneiro sheep comprise a specific group of sheep whose origin stems from the crossbreeding of breeds introduced by the colonizers of Brazil shortly after its discovery. Over time, these sheep have adapted to the unique conditions of the Pantanal region due to its location in a floodplain, characterized by a distinctive ecosystem with diverse flora and fauna. In Brazil, these sheep are found on isolated farms in the region, where there is minimal reproductive control, meaning these sheep essentially live under natural selection processes [1–4].

Adapting these animals to their environment has resulted in unique characteristics compared to other sheep species. One such feature is the distribution of wool on their bodies, whereby little to no wool is observed on their legs, belly, and neck since these areas are prone to prolonged exposure to moisture. Therefore, the presence of wool in these areas would hinder the movement of these animals in the typical waterlogged and densely vegetated areas of the Pantanal. The size of these sheep ranges from small to medium, influenced by the limited food availability. However, these sheep have low caloric requirements and, thus, exhibit the impression of always being lean and accumulating minimal body fat. Fig. 1 (Ref. [5]) presents an image of this sheep breed to illustrate these aforementioned physical characteristics [6,7].

Due to these unique physical characteristics, Pantaneiro sheep are promising subjects for genetic studies, particularly in genomics, which involves the study of genes and the complete genetic material of an organism, encompassing sequencing, mapping, and functional analysis of genes. These various analyses make genomics a vast and critical field with significant impacts across multiple areas of knowledge [8].

The use of DNA genetic markers is common in genomic analyses. These markers refer to a gene or DNA segment with a known chromosome location associated with a specific gene or trait. The marker typically used in the present study is single nucleotide polymorphisms (SNPs), which are molecular DNA markers used to identify mutations and polymorphisms based on the single nucleotide position. Moreover, SNPs are generated through single base substitutions or insertion and deletion events, meaning SNPs act as a source of genetic variation [9–11].

The frequencies of gene loci are systematically influenced by mutations, selection, and migration and are dispersed by natural selection and genetic drift. Hence, understanding which areas in a breed's genome or the specific genes are subjected to selection, particularly natural selection and occasional evolutionary processes, is essential for productive and conservation systems. This knowledge enables the role of adaptation in the evolutionary history of species to be quantified [12,13].

Genetic analysis of animal populations is a scientific approach aimed at understanding genetic diversity, population structure, gene flow patterns, and other genetic aspects within a specific animal species. This type of study is crucial for conserving species biodiversity, sustainably managing wild populations, understanding the evolution of the species being studied, and identifying subpopulations and gene exchanges between different populations [14].

The primary objective of this study was to demonstrate the diversity within the sampled sheep population by identifying the main differences between the Pantaneiro breed and the commercial Texel breed. Another important goal was to associate regions of the sheep genome with the potential effects of the adaptation process and to identify candidate genes subject to the influence of natural selection. A genomic analysis involving DNA genetic markers, specifically SNPs, was performed to achieve these goals.

2. Materials and Methods

2.1 Ethics Statement

All procedures involving animals in this study comply with Animal Use Ethics Committee (CEUA) process number 001/2022, as documented in a certificate signed by the president of CEUA Pantanal. The authorization is valid from April 1, 2022, to December 31, 2024.

2.2 Sampling of Individuals

For this study, samples from 192 adult individuals collected over 4 years of age were divided into seven populations based on their locations: 40 samples from Embrapa (NHFP), 18 from private farms (PRFP), 45 from Dourados (DOSH), 15 from Campo Grande (CGSH), 10 from Aquidauana (AQSH), 19 from Terenos (TESH), and 45 from the commercial Texel herd (TEX). From this division, it can be observed that 147 individuals are Pantaneiro sheep, and 45 belong to the Texel breed. This distribution is detailed in Table 1.

For better visualization of the herd locations, a map of Mato Grosso do Sul was created, showing the location of each farm from which samples for the study were collected. This map presented in Fig. 2 was generated using quantum geographic information system 3.6 (qGIS 3.6, https://qgis.org/).

2.3 Sample Collection

Individual wool samples were collected for DNA extraction. These samples were taken directly from the animals, ensuring the hair follicles were preserved in most collected wool.

2.4 Genotyping

The Embrapa Research Institute performed the genotyping, which was conducted using the GeneSeek® Genomic Profiler™ (GGP) protocol from Neogen (Lansing, MI, USA), utilizing the GGP50K sheep chip. This approach

Fig. 1. Picture of adult Pantaneiro sheep. In this photo, one can observe the distinctive physical characteristics of Pantaneiro sheep: a medium-sized animal with a lean appearance and the absence of wool on the neck, belly, and legs. Reproduced with permission from ref [5].

involves multi-objective localized optimization (MOLO), which focuses on capturing the most relevant and impactful SNPs from previous genotyping matrices. The included SNP content comprised 15,000 SNPs from the GGP15K sheep chip, over 10,000 SNPs from the existing Illumina 50K matrix, and 12,000 SNPs from the Illumina HD content, with over 1700 SNPs added to fill spatial gaps in the genome. Additionally, the GGP Ovine 50K integrates 9000 SNPs derived from a newly sequenced region previously not included in earlier genotyping products. The genotyping procedure adhered to the manufacturer's established guidelines [15].

Plink version 1.9 developed by Shaun Purcell in Boston, MA, USA, was used to assess the quality of the genotypic data [16]. Only SNPs located on autosomal chromosomes were selected. Samples with a call rate below 95% and a minor allele frequency (MAF) less than 0.01 were excluded from the analysis, provided they conformed to the Hardy–Weinberg equilibrium, which was evaluated using Fisher's exact test with 1000 permutations. A p-value threshold greater than 0.05 was considered statistically significant [17].

Following the application of filtering and quality control criteria to the populations, samples meeting the following criteria were removed: call rate below 95%, indicating that the genotyping panel determined less than 90% of genotypes; heterozygosity exceeding three standard deviations from the mean; nearly identical genotypes (over 99.5% similarity); sex identification errors, specifically in individuals listed as males with heterozygous genotypes for markers on the X chromosome. After completing the quality control process, 47,908 SNPs remained [18,19].

2.5 Detection of Outlier SNPs

Outliers formulate observations or data points in a dataset that stand out significantly from the general pattern. In statistics and data science, values are substantially distant from the mean or central tendency of the data. In genetics, outlier SNPs refer to specific genetic variants that differ significantly from expected based on the general genetic distribution within a population. These SNPs may indicate something unusual, such as an evolutionary adaptation or a response to a specific environmental condition [20].

Three distinct methods were employed to detect outlier SNPs, as detailed below. Indeed, applying multiple methods aims to broaden the range of results, increasing the likelihood of identifying genes of interest.

The first method utilized was pcadapt, implemented in version 2.0 of the R package. Subsequently, pcadapt was used to identify outlier SNPs through exploratory genomic

Table 1. Location, nomenclature, and quantity of samples from Pantaneiro sheep.

Acronym	Number of individuals	Location	Farm coordinates	
NHFP	40	Nhumirim	18°59′17.11″S	56°37′8.59″O
PRFP	18	Santa Glória Farm	17°52′33.92″S	55°50′27.57″O
		San Saruê Farm	17°33′47.62″S	56°7′48.74″O
		Viveirinho Farm	17°46′39.00″S	56°5′32.16″O
		Imaculada Farm	19°15′0.57″S	57°1′23.28″O
		Figueiral Farm	18°23′48.39″S	55°24′44.61″O
		Santa Mônica Farm	17°37′55.07″S	56°21′23.99″O
		Carvalho Farm	17°34′4.78″S	56°29′15.20″O
DOSH	45	UFGD Farm	22°11′49.89″S	54°56′22.47″O
CGSH	15	Uniderp Três Barras Farm	20°33′52.44″S	54°32′29.18″O
AQSH	10	UEMS Aquidauana	20°27′27.37″S	55°40′7.93″O
TESH	19	CNPGC Terenos Farm	20°33′10.34″S	54°48′43.89″O
TEX	45	Texel Farm	21°29′11.92″S	57°28′19.41″O

NHFP, Embrapa; PRFP, private farms; DOSH, Dourados; CGSH, Campo Grande; AQSH, Aquidauana; TESH, Terenos; TEX, Texel; UFGD, Federal University of Grande Dourados; UEMS, State University of Mato Grosso do Sul; CNPGC, National Beef Cattle Research Center.

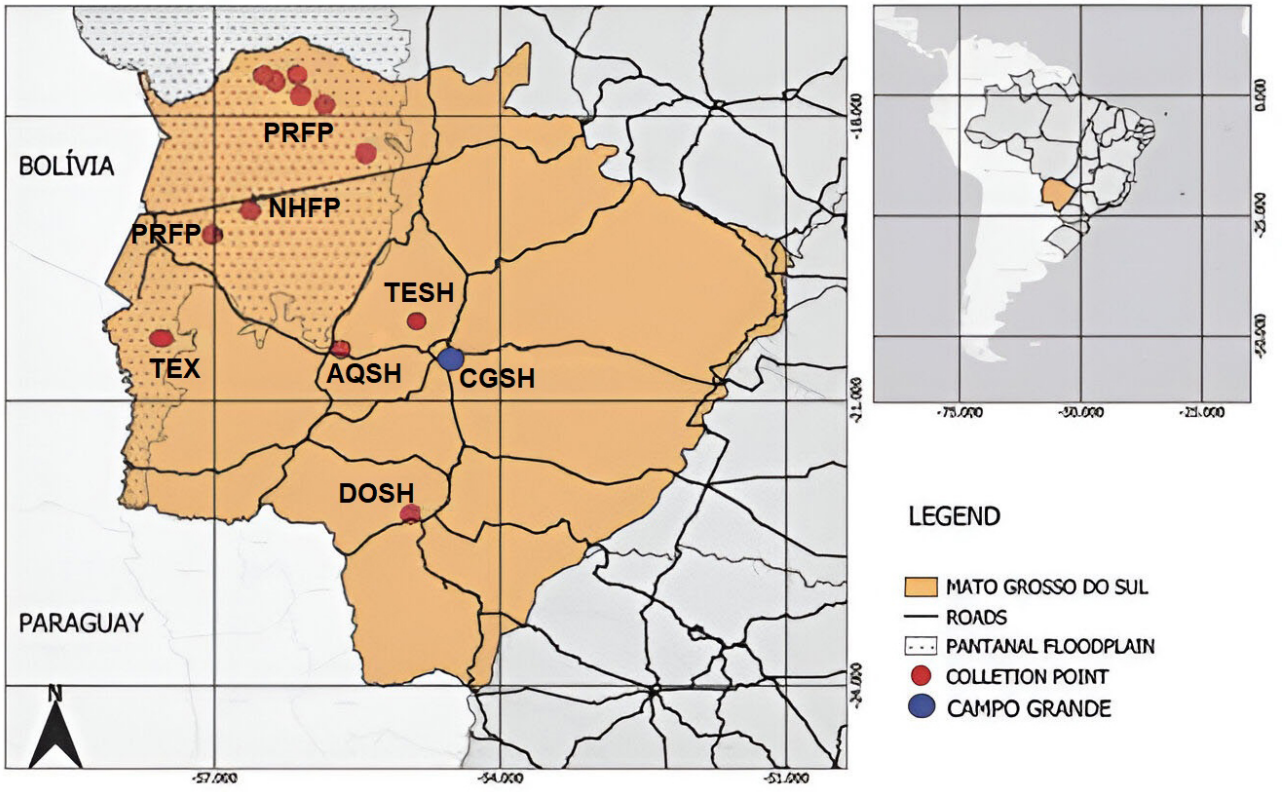

Fig. 2. Map of Mato Grosso do Sul, indicating the location of each herd. In orange is the state of Mato Grosso do Sul, the places with small dots, and where the Pantanal is located in the state. The blue dots and the red dots indicate the farm locations that contain the herds of Pantaneiro sheep.

analyses, including principal component analysis (PCA). Initially, a single principal component (K = 1) was used, as it explained most of the variation. Then, Mahalanobis distances were employed as a statistical test to identify outlier SNPs, which were subjected to hypothesis testing and converted to *p*-values. A q-value threshold of 0.01 was applied to control for the false discovery rate [21] and determine the atypical loci.

The second method used was OutFLANK, which was again implemented in R. This method employs the proba-

bility of a truncated fixation index (F_{ST}) distribution to infer the distribution related to neutral markers. This distribution assigns *p*-values to each locus, allowing the identification of outliers that may indicate heterogeneous selection. R packages devtools 2.4.5, OutFLANK 0.2 , and vcfR 1.15 were used to facilitate the graphical visualization of SNPs with significant differences [22].

The third and final method was based on the approach presented in 1996 by Beaumont and Nichols [23]. The FDIST2 method was used to obtain the null distribution of F_{ST} compared to expected heterozygosity under an isolation scenario. The fsthet R package 1.0.1 was employed to compute smoothed quantiles from genotypic data, identifying outlier loci with extreme F_{ST} values relative to Hardy–Weinberg equilibrium heterozygosity. These outliers are considered to have potential adaptive significance [23–27].

2.5.1 False Discovery Rate (FDR)

In conceptual terms, the false positive rate represents the number of significant neutral loci (false positives) divided by the total number of neutral loci tested. Similarly, the false discovery rate (FDR) can be defined as the number of false-positive neutral loci divided by the total number of positive results. These results were converted into q-values to adjust for multiple comparisons. This correction was applied only to the pcadapt and OutFLANK methods [28].

A threshold of q = 0.01 was adopted to identify a positive result. The q-value of a locus represents the estimated proportion of false positives among all loci with *p*-values equal to or less than that of the locus in question. Under this premise, loci with q-values of 0.01 are associated with an expected FDR of 1%; the q-values method from the R package was applied, which converts *p*-values into q-values [21,29].

2.5.2 Gene Content of Regions under Natural Selection

The National Center for Biotechnology Information (NCBI) Genome Data Viewer tool was utilized to locate the genes within the significant regions identified as under selection, specifically searching for the organism *Ovis aries*. The reference sheep genome used was the most currently available in the tool, named ARS-UI_Ramb_v3.0 (https://www.ncbi.nlm.nih.gov/gdv/browser/genome/?id=GCF_016772045.2). For this purpose, all outlier SNPs identified by the three detection methods were compiled, and the position of each SNP on the corresponding chromosome was entered into the aforementioned search tool. This allowed for genes within these regions to be identified. Then, all identified genes were annotated, and their functions were researched in the literature. Genes with phenotypic functions were specifically highlighted as results in this study.

3. Results

3.1 Population Diversity Analysis

Table 2 presents the parameters used to describe various aspects of genetic variability in populations as applied in population genetics studies. These parameters are calculated based on the allele distribution within a sample of individuals. The value of "Ar" represents the average number of alleles present at a specific locus within the population; thus, a higher "Ar" value indicates greater allelic diversity [30].

Table 2 also shows additional parameters, such as expected heterozygosity (He), which refers to a measure of the expected genetic diversity within a population. The He parameter represents the probability that a randomly chosen individual will be heterozygous at a specific locus. Conversely, the observed heterozygosity (Ho) measures the actual genetic diversity observed within the population, indicating the frequency of heterozygotes relative to the total number of individuals examined at a given locus [31].

The level of inbreeding within a population is indicated by the inbreeding coefficient (F_{IS}). A positive F_{IS} value suggests a tendency towards inbreeding, whereas a negative value indicates higher heterozygosity than expected. The last two columns in Table 2 show the lower and upper confidence intervals for F_{IS}, calculated using bootstrap techniques [32–34].

Table 2 reveals a distinction between the Pantaneiro and the commercial Texel breeds. Notably, the Texel herd exhibited the highest expected heterozygosity and the low-

Table 2. Diversity among the seven populations based on allele distribution.

Population	n	Ar	He	Ho	F_{IS}	Fis.bs.fis.CI.ll	Fis.bs.fis.CI.hl
NHFP	40	1.998	0.434	0.398	0.095	0.093	0.096
PRFP	18	1.998	0.436	0.393	0.129	0.127	0.131
DOSH	45	1.998	0.435	0.414	0.059	0.058	0.060
CGSH	15	1.998	0.431	0.411	0.080	0.077	0.081
TESH	19	1.998	0.434	0.412	0.079	0.077	0.082
AQSH	10	1.998	0.417	0.419	0.048	0.045	0.051
TEX	45	1.999	0.448	0.385	0.151	0.150	0.153

Fis.bs.fis.CI.ll would be the lower limit of the confidence interval for the Fis estimate, and Fis.bs.fis.CI.hl would be the upper limit. He, expected heterozygosity; Ho, observed heterozygosity; F_{IS}, inbreeding coefficient.

est observed heterozygosity, alongside a higher F_{IS} value, suggesting a greater tendency towards inbreeding than the Pantaneiro breed.

Fig. 3 presents a visual representation of the genetic interactions among the sheep populations through a network graph, providing valuable insights into the structure and connectivity of these communities. A distinct pattern is observed in this context: six of the seven populations show a notable proximity on the network map, while the Texel population is conspicuously more isolated.

The sheep population network graphically illustrates the genetic relationships among different groups, with individuals represented as nodes and their connections as lines indicating genetic proximity.

In this case, the clustering of six populations suggests a more intense genetic interaction characterized by allele sharing and common ancestry, indicating a more intertwined genetic history. In contrast, the Texel population stands out due to its more isolated position on the network graph. This genetic distance may be attributed to various factors, such as the reduced impact of natural selection, geographical isolation, or a unique set of genetic characteristics that differentiate it from other populations.

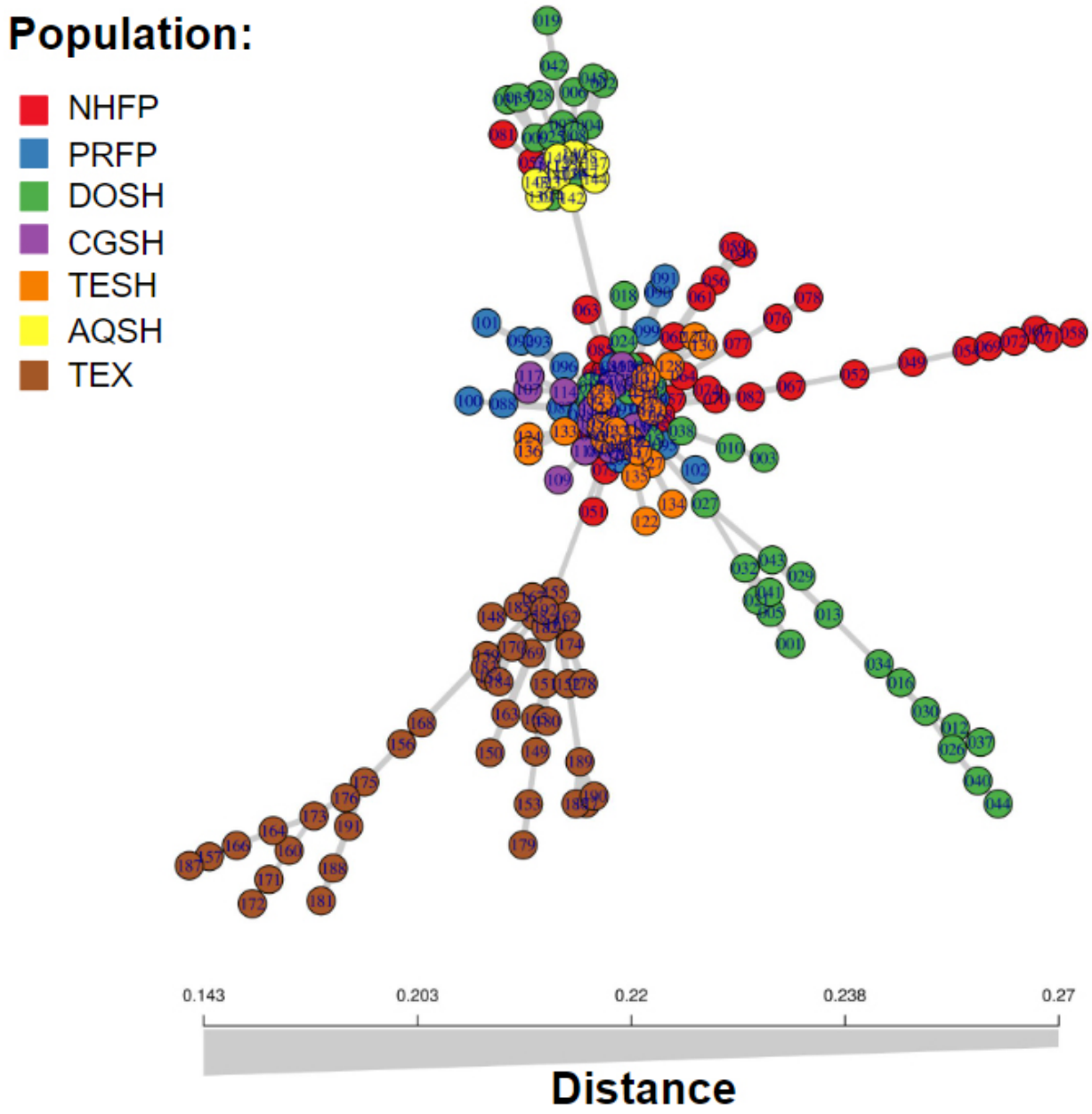

Fig. 3. Network graphical illustration of the distances between the seven populations. The colored dots represent each Pantanal sheep; each color represents a different population, as indicated in the legend; genetically closer individuals are located nearer to each other.

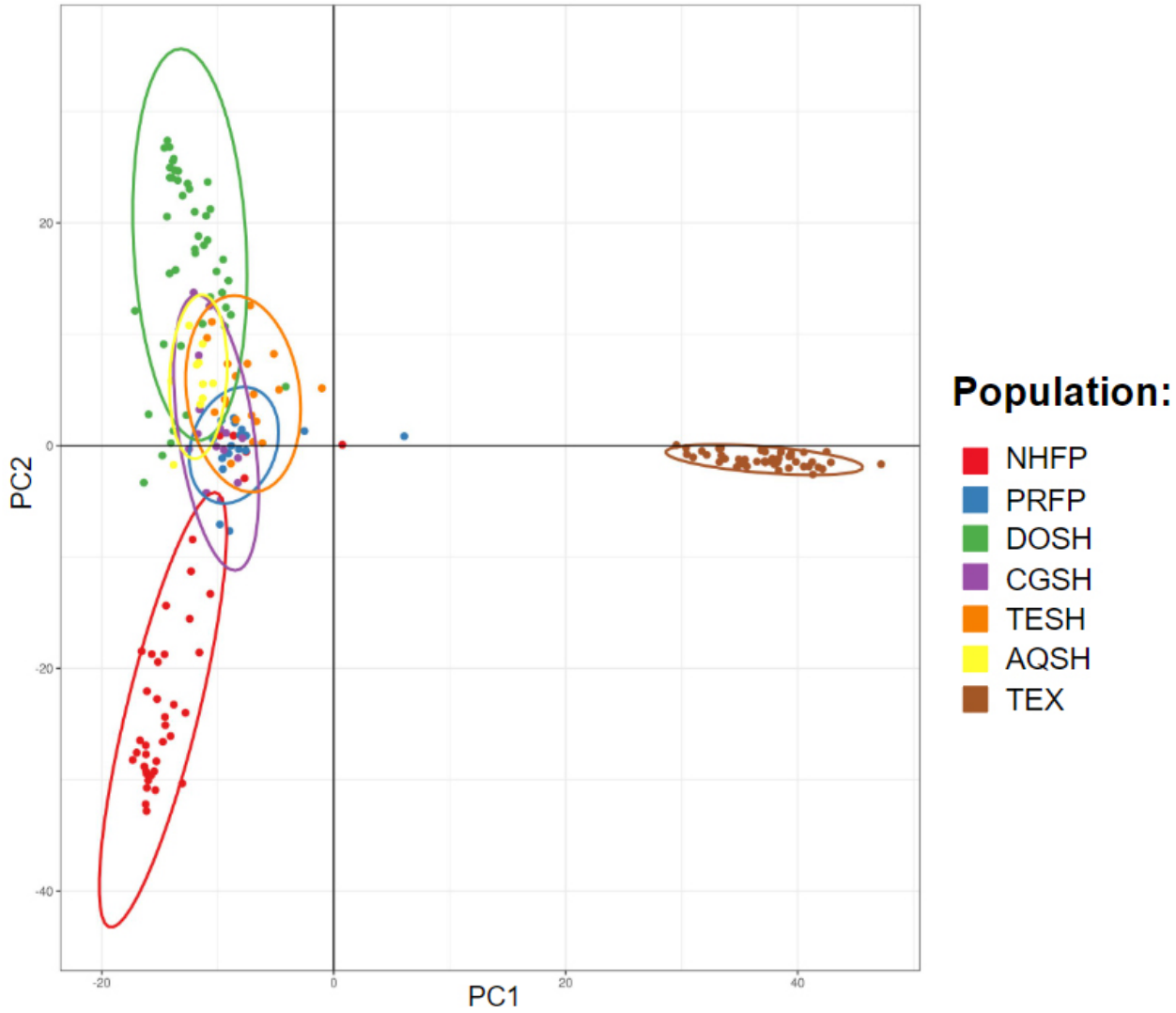

Fig. 4. Principal component analysis plot for the seven sheep populations. This graph shows the population clusters found for the seven populations, where each point is an individual, and each color is a population. PC, principal component.

The PCA plot shown in Fig. 4, with Principal Component (PC)1 on the x-axis and PC2 on the y-axis, provides valuable insights into the variability and similarity among sheep populations; this type of plot is commonly used to visualize complex patterns in high-dimensional datasets [35].

In the specific context of this plot, the seven distinct sheep populations are represented by colors, as indicated in the legend. The clustering of six populations towards the left side of the plot suggests a strong similarity among these groups. These six populations are the Pantaneiro sheep with similar characteristics, meaning each is positioned closely in the principal component space. In contrast, the Texel population exhibits unique characteristics that differentiate this breed from the Pantaneiro population.

Thus, the interpretation of this PCA plot suggests a high degree of genetic homogeneity among the six Pantaneiro populations. In contrast, the Texel population stands out as a distinct entity characterized by variable differences. This separation is likely attributable to variations in reproductive control practices across farms.

This visualization plays a crucial role in elucidating the structure of the data, enabling the identification of distinct groups and revealing underlying patterns that warrant further investigation in future research. This comprehensive overview establishes a solid foundation for conducting more nuanced analyses of the genetic traits that distinguish the sheep populations depicted in Fig. 4. This preliminary insight facilitates a deeper understanding of the genetic diversity among the populations and directs future studies toward exploring the specific genetic factors contributing to these observed differences [35].

Fig. 5 presents a dendrogram displaying all 192 analyzed individuals separated into seven populations. As this graph includes every individual, the clearest interpretation is achieved through the color-coding of each population. The Pantaneiro populations exhibit greater diversity due to

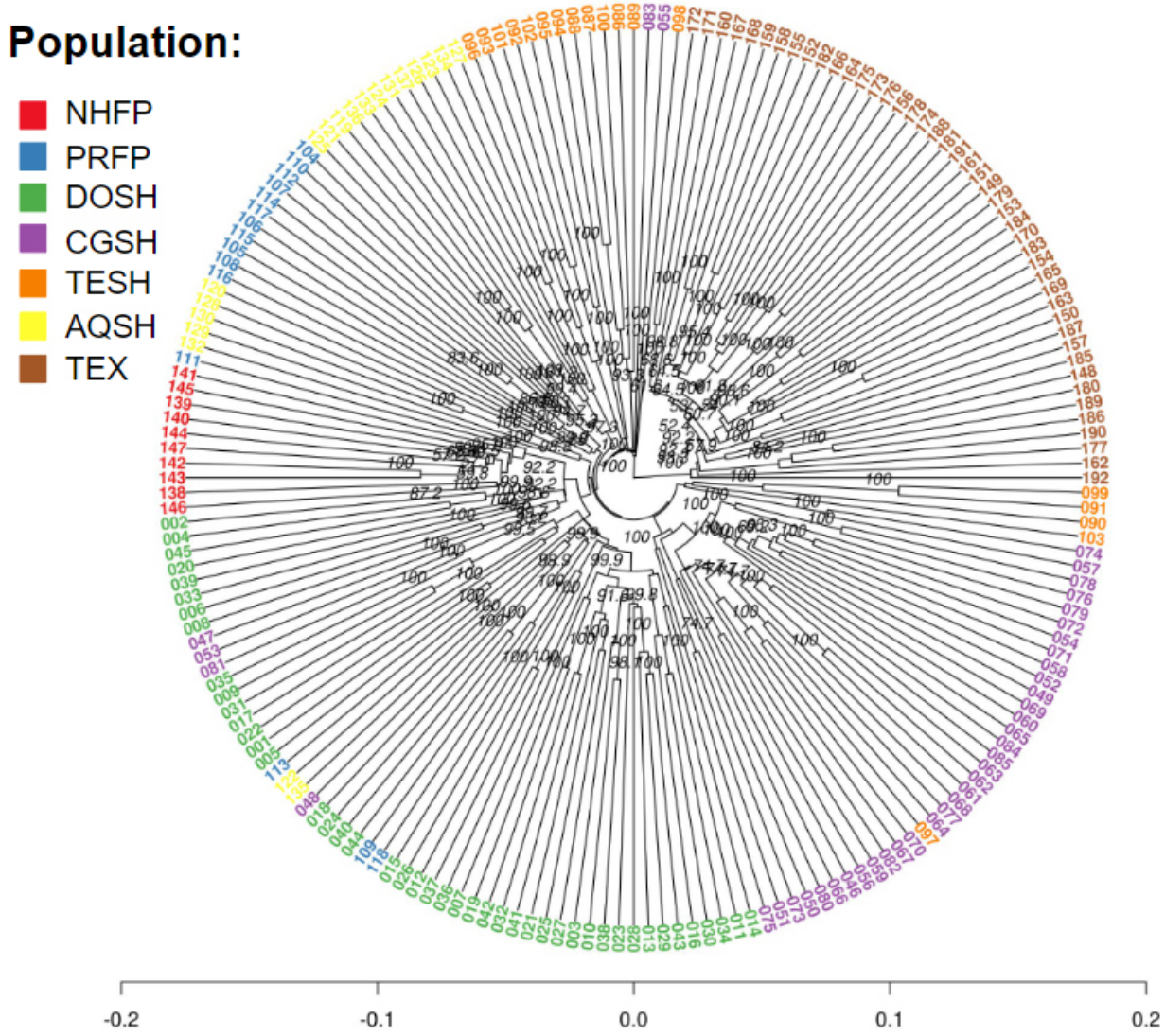

Fig. 5. Circular dendrogram for the seven populations. The numbers around the circle represent the 192 individuals; the colors represent the populations; the lines represent the genetic similarity of each individual.

the lower level of control, whereas the Texel population appears distinctly separate within its section of the dendrogram.

When isolating data for the Texel breed and focusing solely on the Pantaneiro breed populations, it becomes evident that these populations are well-mixed in the graphs, indicating significant genetic variation among them. However, two populations exhibited distinct characteristics: Those from the Nhumirim (NHFP) and Dourados (DOSH) regions. Parts of these populations diverged in different ways in both the PCA and network graphs, suggesting that sheep in these regions may possess unique genetic traits not shared with other populations. This finding is intriguing and warrants further investigation in future studies.

To better visualize the results of each presented graph, refer to Fig. 6 (Ref. [36,37]), where panel A represents the Pantaneiro sheep and panel B represents the Texel sheep. The morphological characteristics of these breeds explain the significant distance between them in the graphs. Texel sheep have denser wool that covers their bellies and appear more robust and physically stronger. In contrast, Pantaneiro sheep have less wool and a leaner appearance [36,37].

Global Fixation Index (Global F_{ST})

F_{ST} values were obtained from a list of intervals termed "F_{ST} Upper" and "F_{ST} Lower" using the R programming language and the relevant R package. A global F_{ST} value of approximately 0.10 was determined by calculating the mean of each interval and the overall mean.

The interpretation of an F_{ST} value ranges from 0 to 1, where 0 indicates no genetic differentiation between populations, with all alleles being equally distributed among them, and 1 signifies high genetic differentiation, where each population possesses unique alleles with little to no

Fig. 6. Morphological differences between the Pantaneiro and Texel breeds. (A) Pantaneiro sheep. (B) Texel sheep. Reproduced with permission from ref [36,37].

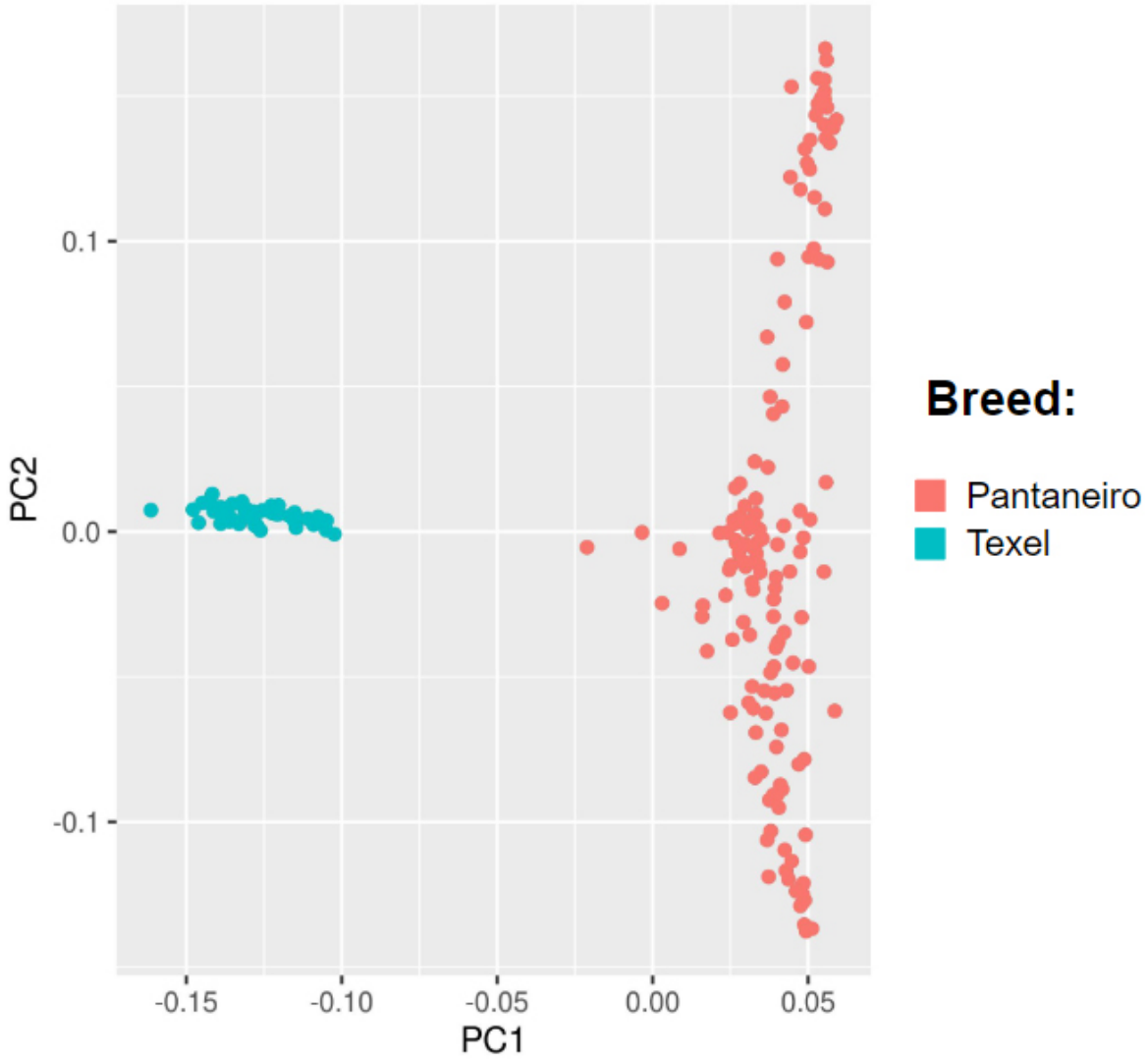

Fig. 7. Clustering individuals into two populations based on principal components (PCs). This graph shows the distribution of two populations, where the Texel is in blue, and the rest, called the Pantanal, is in red. The great difference between the breeds is illustrated.

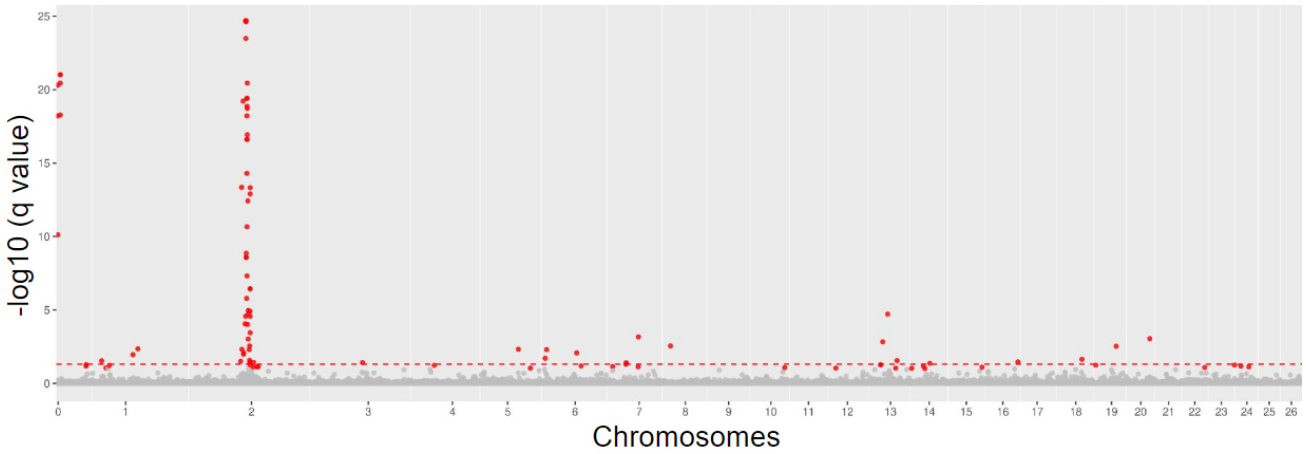

Fig. 8. The q-value for each chromosome using the pcadapt method. The red dots represent the single nucleotide polymorphisms (SNPs), and each column represents a chromosome of the animal's genome above the cut line, meaning they are considered outliers.

genetic exchange. The global F_{ST} value of 0.10 observed in this study reflects moderate genetic differentiation, suggesting that some genetic exchange exists between the populations and notable differences are present in the alleles [38].

This moderate differentiation is attributed to varying levels of control across populations. For example, the Texel sheep exhibit higher control than the Pantaneiro breed, contributing to the observed genetic differentiation.

3.2 Outlier SNPs and Genomic Regions Under Natural Selection

The pcadapt method created a plot visualizing the principal components (PCs) to group individuals genetically, as shown in Fig. 7. This plot clearly illustrates the distinction between the commercial Texel breed and the Pantaneiro breed.

Following the implementation of the pcadapt method, a plot was generated to visualize the q-values for each chromosome, as shown in Fig. 8. The neutrality threshold is indicated at 1.3, adjusted for the FDR, where all points above this dashed line represent outlier loci. The plot reveals that chromosome 2 exhibits a marked difference, with numerous outlier loci. Notably, chromosome "0" indicates the unmapped regions or mitochondrial DNA, against which the program cannot distinguish; thus, only the identified SNPs are presented.

A total of 56 outlier SNPs were detected after applying the FDR directly to the OutFLANK method. Fig. 9 illustrates the relationship between the F_{ST} values and the SNP positions, with outlier loci highlighted in pink. The plot shows that the more significant results are discovered for SNPs around position 8000, corresponding to chromosome 2.

For the FDIST2 method, implemented via the fsthet R software, 292 outlier loci were detected, exhibiting F_{ST} values above the neutrality zone, with a mean of 0.29 and average heterozygosity of 0.45. These values are shown in Fig. 10, which displays the relationship between F_{ST} and total heterozygosity (H_T) values. Outlier loci, highlighted in red, indicate regions under potential inbreeding. This method did not apply the FDR correction to maximize the number of observed outlier SNPs.

Genes under Selection Associated with Phenotypic Traits

To identify genes associated with the outlier SNPs found using the three methods applied in the study, the NCBI tool mentioned in the Materials and Methods section was utilized, referring to the ovine genome ARS-UI_Ramb_v3.0. This approach yielded 15 characterized genes using the pcadapt method, 15 with the OutFLANK method, and 145 with the FDIST2 method. The high number of genes identified using the FDIST2 method was attributed to the absence of the FDR correction.

After collating all the identified genes and researching their functions in the literature, several genes of particular interest were found to be responsible for the visible traits in the animals. Identifying genes of this nature is crucial for understanding desirable phenotypic traits under selection. Table 3 lists the identified genes, the method used to recognize the SNP, chromosome, and respective positions.

Notably, numerous genes encoding proteins involved in specific processes were found; however, due to the high number and their specificity, only those associated with phenotypic factors, such as physical appearance, behavior, and physiological traits, are presented here.

A literature review was conducted for each gene to verify the gene function and assess their phenotypic roles. The following functions were observed:

FGF12, located on chromosome 1 and identified using the fsthet method, is part of the fibroblast growth factor (FGF) family. According to Woo, Suh, and Sung [39], the *FGF12* gene family functions in biological processes such as cell growth, tissue regeneration, embry-

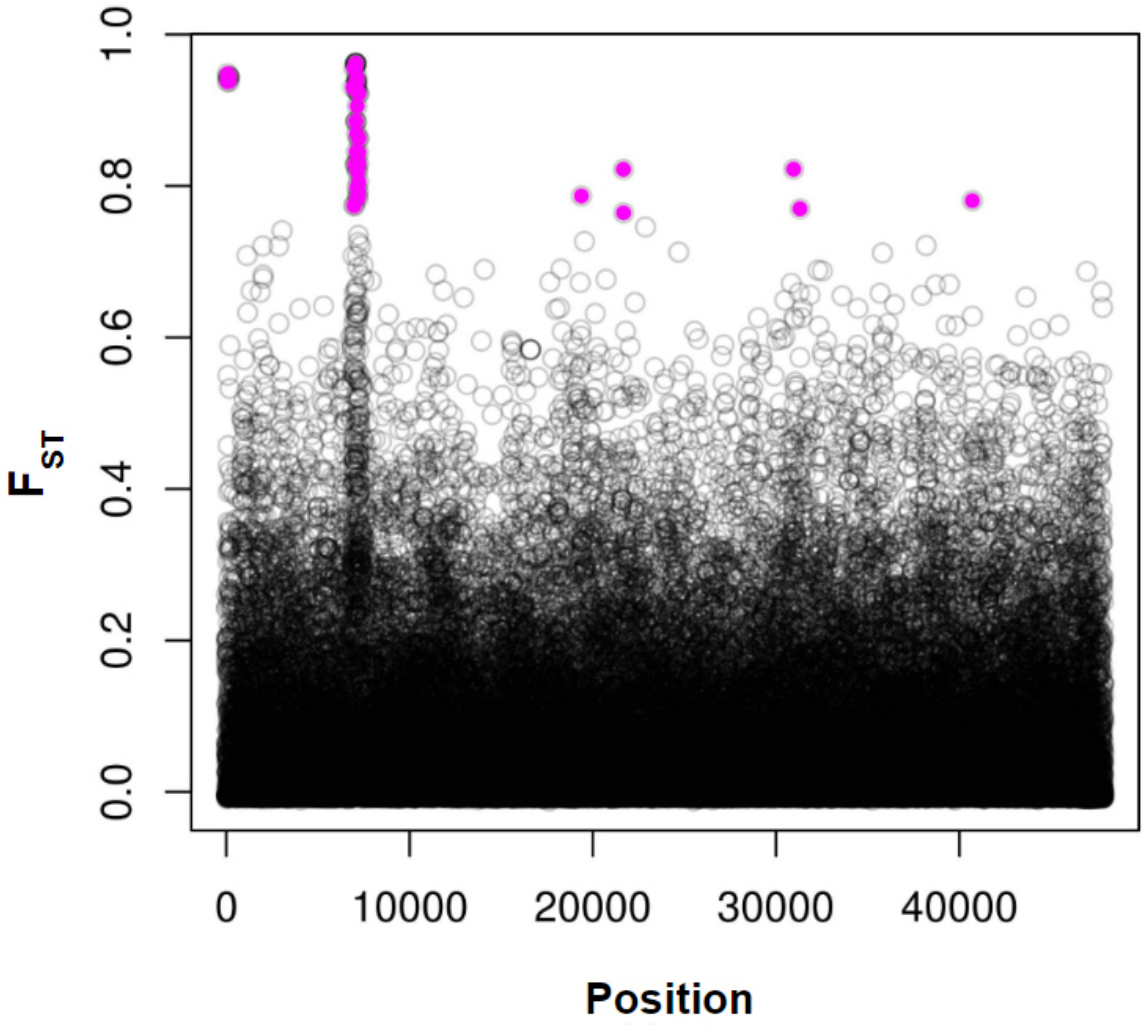

Fig. 9. Relationship between outlier position and Fixation Index (F_{ST}) value for the OutFLANK method. Each point represents a SNP, and outlier SNPs are colored purple; in this graph, it is possible to observe the SNP positions in the genome.

onic development, and angiogenesis. Notably, the study by these researchers highlighted that *FGF12* contributes to hair growth, suggesting that the *FGF12* gene affects wool growth in sheep [39].

In the pcadapt and OutFLANK methods, *SCRG1* was observed on chromosome 2 with the same SNP, indicating overlap. This gene is associated with aiding tissue and bone regeneration and regulating chondrogenesis. It also produces a protein in articular cartilage [40].

Moreover, there was an overlap between the pcadapt and OutFLANK methods, with *OCA2* appearing on chromosome 2 at adjacent SNPs. According to Loftus *et al.* [41], *OCA2* is responsible for pigmentation in mammals, influencing skin color variation and acting as a determinant of eye color.

Finally, *TPM1*, located on chromosome 7 and identified using the pcadapt method, plays a role in the contractile system of striated and smooth muscles and the cytoskeleton of non-muscle cells [42].

Observing the presence of such genes is crucial for understanding the physical traits under selection in these animals, such as changes in coat texture, skin, and eye pigmentation, or even cellular, tissue, and muscle development.

4. Discussion

4.1 Population Diversity Analysis

The analysis of genetic variability parameters between sheep populations revealed significant differences between the Pantaneiro breed and the commercial Texel breed. While the Texel breed exhibits increased expected heterozygosity and lower observed heterozygosity, suggesting a tendency towards inbreeding, the Pantaneiro breeds demonstrate broader genetic diversity. This is evidenced by the inbreeding coefficient (F_{IS}), which is higher for Texel sheep, indicating greater consanguinity within this group. This observation aligns with the fact that the Texel breed is managed under more stringent reproductive control, whereas Pantaneiro sheep are either maintained with

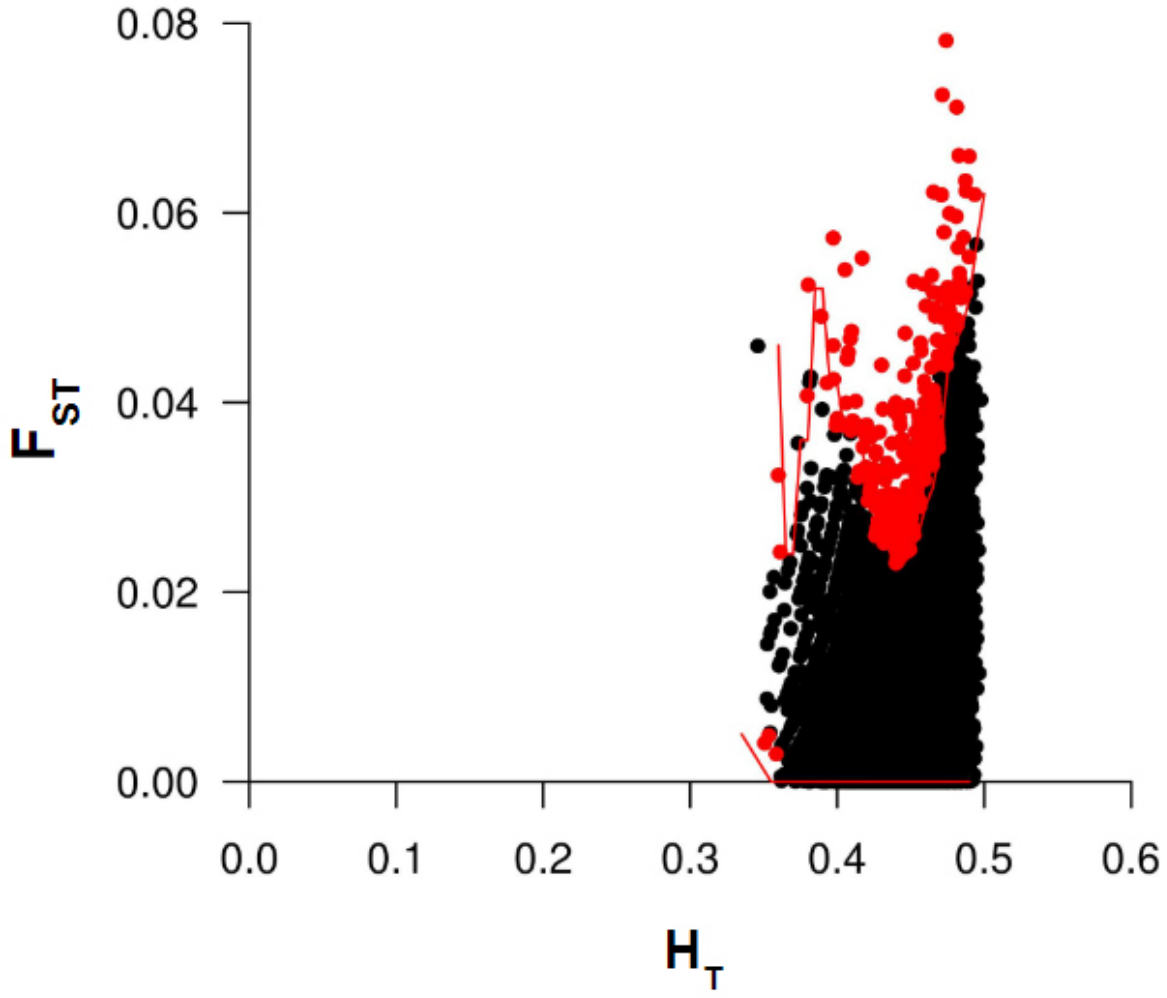

Fig. 10. Relationship between F_{ST} value and total heterozygosity (H_T) for outliers detected using the fsthet method. Each point represents a SNP; those colored red indicate outliers, and the large number of outliers suggests the lack of a false discovery rate (FDR) correction.

Table 3. Genes found in outlier SNPs that transcribe phenotypic characteristics.

Method	SNP	Chromosome	Position	Gene (NCBI)
Fsthet	3603	1	195943567	*FGF12* (fibroblast growth factor 12)
pcadapt OutFlank	6905	2	107288479	*SCRG1* (chondrogenesis stimulator 1)
pcadapt OutFlank	7034 7033	2	113546352	*OCA2* (*OCA2* melanosomal transmembrane protein)
pcadapt	21678	7	45184872	*TPM1* (tropomyosin 1)

strategies to preserve genetic diversity or are left under natural selection in breeder-owned flocks, resulting in greater genetic sharing.

Visual representations of genetic interactions through network and PCA graphs show that Pantaneiro breeds are more genetically interconnected, indicating a shared genetic history. In contrast, the Texel breed appears more isolated in the graphs, suggesting a significant genetic distinction from the Pantaneiro breeds.

The global fixation coefficient (F_{ST}) calculated from the data also indicates moderate genetic differentiation between the populations, with a value of approximately 0.10. This suggests that some gene exchange occurs among populations and that there are notable differences in allele frequencies, which can be partially attributed to the varying levels of reproductive control between the Pantaneiro and Texel breeds. This value indicates a sufficient level of distinction and suggests some degree of population structure,

which may be relevant for conservation studies, genetic selection, or local adaptation.

Crispim *et al*. [43] reported similar results using a different methodology, specifically microsatellite markers when comparing seven different breeds. Indeed, Crispim and co-authors observed greater allele richness and higher genetic diversity in the Pantaneiro breed compared to the Texel breed. Consequently, the authors highlighted the substantial genetic differentiation of Pantaneiro sheep from other breeds, suggesting that Pantaneiro sheep may serve as a valuable reservoir of genetic diversity. This diversity may include rare and economically important alleles, making the Pantaneiro breed potentially useful in breeding and genetic management programs [43].

Other studies have also utilized SNPs in sheep, such as the work by McManus *et al*. [44], which explored spatial patterns of genetic structure across various sheep breeds in Brazil using SNPs to examine relationships between these breeds. An interesting finding relevant to the current study is that the Pantaneiro breed from the Central–West region may have originated from the Crioula wool breed. However, it also exhibits a genetic composition influenced by crossbreeding, incorporating alleles from wool sheep breeds such as Bergamácia and hair breeds from the Northeast region of Brazil [44].

Oliveira *et al*. [45] examined several genetic parameters in Pantaneiro sheep, with one finding highlighting the need for caution regarding inbreeding. Oliveira *et al*. noted that increased genetic contribution could elevate kinship relationships, increase inbreeding, and reduce the effective population size. Therefore, controlling mating in this flock is recommended to prevent inbreeding depression. However, in the current study, only the Texel breed showed significant effects related to inbreeding, as indicated by the data [45].

The study conducted by Embrapa Pantanal, as reported by Oliveira *et al*. [37], on the morphology of Texel sheep demonstrates that no genes from other breeds were introduced into this flock through the homogeneity of morphological variable results. Consequently, these sheep are eligible for genealogical registration with the Brazilian Association of Texel Breeders due to the minimal variability in the studied phenotypic characteristics. This finding aligns with the results of this study, as the Texel breed was distinctively separated from the other sheep, showing high inbreeding and low genetic variation [37].

Comparing results across different studies and breeds is essential for a more comprehensive understanding of genetic diversity and adaptation. Such comparisons validate findings within the specific context of each study and highlight broader patterns and trends that may not be evident in isolated analyses.

4.2 Identification of Outlier SNPs

The results obtained using the three methods for detecting outlier SNPs in Pantaneiro sheep provide valuable insights into the genetic structure of these animals and the selection processes potentially occurring within the studied populations.

The application of the PCA clearly demonstrated the genetic distinction between the commercial Texel breed and the Pantaneiro sheep, underscoring the utility of this technique for identifying genetic clusters.

The analysis of genes located in these outlier regions revealed the presence of several well-characterized genes, as shown in Table 3, associated with visible phenotypic traits in Pantaneiro sheep. These genes influence hair growth, tissue regeneration, pigmentation control, and muscular capacity, highlighting the significance of genetic variation in the adaptation of these animals and guiding conservation and management strategies.

Given the scarcity of studies utilizing SNP markers in Pantaneiro sheep, the work by Garza Hernandez *et al*. [46] was particularly relevant. This study, which uses SNP markers in Texel sheep with similar genotyping methodologies, identified important genes associated with muscle weight, body and carcass weights, parasite resistance, and transcriptional regulation, among other traits. Identifying a similar study is crucial to demonstrate that finding genes associated with SNPs in these animals is possible and closely aligns with the present study [46].

Another relevant study was conducted by Kumar *et al*. [47], which focused on conservation genetics using a similar approach. This study employed a 50K sheep SNP chip and applied a different methodology, MAF-LD, to identify specific SNPs in three Indian sheep breeds: Garole, Deccani, and Changthangi. Although these breeds differ from those discussed in the present article, the study yielded interesting and comparable results. Kumar *et al*. [47] discovered that these SNPs were associated with production and disease resistance-related genes. Subsequently, this methodology could be practically applied to determine whether individuals are of purebred status [47].

Comparing studies that utilize SNPs in sheep provides critical insights into the consistency and variability of genetic findings across different contexts. Such comparisons allow for assessing similarities and differences using genetic markers related to areas of interest. By examining the results of other studies, researchers can gain a deeper understanding of various SNP-based approaches. This comparative analysis not only strengthens the reliability of the findings but also informs future research directions and guides applications in breeding and conservation.

5. Conclusion

Throughout the study, it became evident that sheep farming has experienced significant growth and has established its importance for both livestock producers and re-

searchers. Genetics has proven to be a valuable tool in studying these animals, as it provided significant insights into the genetic structure and diversity of Pantaneiro sheep populations, with important implications for understanding their adaptation and biodiversity conservation.

Identifying outlier SNPs revealed patterns of genetic selection across different genomic regions, highlighting the influence of evolutionary processes and the genetic diversification of the studied populations. The various employed methods offered valuable and consistent insights into genetic variability and interactions between populations. Notably, genes associated with phenotypic traits such as hair growth, tissue regeneration, pigmentation control, and muscular capacity were identified.

Furthermore, the assessment of population diversity through genetic parameters revealed marked differences between the Pantaneiro breed and the commercial Texel breed. The Pantaneiro breed exhibited greater genetic diversity, while the Texel, a more controlled herd, showed expected characteristics of high inbreeding and low genetic exchange. These findings are crucial for a better understanding the evolutionary history and genetic interactions among sheep populations in the Pantanal region of Mato Grosso do Sul. Moreover, the findings highlight the importance of preserving genetic diversity and adapting species to ensure their continued viability.

Availability of Data and Materials

All data points generated or analyzed during this study are included in this article and there are no further underlying data necessary to reproduce the results.

Author Contributions

LS wrote and edit the manuscript, made the editorial changes and wrote the original draft preparation. LS, MB, FDS, AA and FVJ curated and analyzed data. MB, AA and FVJ started the project, partnered and administered. MB and FDS did the programming part. AA and FVJ administered and did the funding acquisition. FVJ: conceptualized the study, validated, resourced. FVJ, MB, MF and FR developed the methodology. FVJ, GVB and FR carried the investigation. FVJ, FR, MF, MB, AA and FDS helped with writing, review and editing. FVJ, MB and AA supervised. RJ: Collected the samples. All authors viewed, read, reviewed and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

All procedures performed with animals in this study are in accordance with CEUA process number 001/2022, through a document signed by the president of CEUA Pantanal, whose authorization is valid from April 1st, 2022 to December 31, 2024.

Acknowledgment

We would like to express our gratitude to the State University of Santa Catarina and Embrapa Pantanal for their support and resources, which were crucial to the success of this research and made it all possible. We also extend our thanks to the Alma Pantaneira Institute (IAP) and Médicos do Pantanal (MDP) for their assistance in collecting samples from Pantaneiro sheep, and to veterinarian Elton Bock for his support in collecting samples from Texel animals.

Funding

The authors declare that financial support was received for the research and development of this article. This research was funded by Project FUNDECT No. 355/2022, SIAFEM No. 32366. The approved project was titled: "Conserving, developing, and consolidating the Pantanal sheep herd as the first genuinely Mato Grosso do Sul breed."

Conflict of Interest

The authors declare that there is no conflict of interest with Embrapa or any other company regarding the research and publication of this manuscript. Financial support for this work was provided through a project funded by FUNDECT, and this funding does not give rise to any conflict of interest with Embrapa or any other company.

References

[1] Juliano RS, Egito AA, Araujo FAS, Ruiz MA. Pantanal Bovine: The Best of the Pantanal Translated into Livestock, Culture, Tradition, and Biodiversity. RG News Journal. 2022; 8: 22–28. (In Portuguese)

[2] Mariante AS, Albuquerque MSM, Egito AA, McManus C. Advances in the Brazilian animal genetic resources conservation programme. Animal Genetic Resources Information. 1999; 25: 107–121.

[3] Oliveira DP, Oliveira CAL, Nunes ME, Vargas Junior FM, Ferreira MB, Seno LO, *et al*. Morphostructural Characterization of Young Female and Male Pantanal Sheep, a Native Breed from Mato Grosso do Sul, Brazil. Semina: Agricultural Sciences. 2014; 35: 973–986.

[4] Silva JSV, Abdon MM. Delimitation of the Brazilian Pantanal and its sub-regions. Brazilian Agricultural Research. 1998; 33: 1703–1711. (In Portuguese)

[5] Santos AS, Juliano RS, Paiva SR, Araújo MTBD, Berselli C. Description of traditional sheep farming systems in Nhecolândia, Pantanal, MS. Circular Técnica. 2010; 94: 1–5. (In Portuguese)

[6] Costa JAA, Egito AA, Barbosa-Ferreira M, Reis FA, Vargas Junior FM, Santos AS, *et al*. Pantanal sheep, a genetically adapted group in the state of Mato Grosso do Sul, Brazil. Embrapa. 2013; 8: 25–43. (In Portuguese)

[7] Vargas Junior FM, Martins CF, Souza CC, Pinto GS, Pereira HF, Camilo FR, *et al*. Avaliação Biométrica de Cordeiros Pantaneiros. Agrarian Journal. 2011; 4: 60–62. (In Portuguese)

[8] Garcia ES. Genes: Facts and fantasies. Fiocruz: Rio de Janeiro. 2006. (In Portuguese)

[9] Al-samarai FR, Al-kazaz AA. Molecular markers: an introduction and applications. European Journal of Molecular Biotechnology. 2015; 9: 118–130.

[10] Caetano AR. SNP markers: basic concepts, applications in animal management and breeding, and perspectives for the future. Brazilian Journal of Animal Science. 2009; 38: 64–71. (In Portuguese)
[11] Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, *et al*. Genome-wide mapping with biallelic markers in Arabidopsis thaliana. Nature Genetics. 1999; 23: 203–207.
[12] Brito LF, Kijas JW, Ventura RV, Sargolzaei M, Porto-Neto LR, Cánovas A, *et al*. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC Genomics. 2017; 18: 229.
[13] Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics. 1973; 74: 175–195.
[14] Hartl DL, Clark AG. Principles of population Genetics. 4th edn. Sinauer Associates: Oxford. 2006.
[15] Hofeneder-Barclay H. NEOGEN Genomics: GGP Ovine 50K. Neogen. 2021; 1: 4–20.
[16] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015; 4: 7.
[17] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics. 2007; 81: 559–575.
[18] Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nature Protocols. 2010; 5: 1564–1573.
[19] Kijas JW, Ortiz JS, McCulloch R, James A, Brice B, Swain B, *et al*. Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. Animal Genetics. 2013; 44: 325–335.
[20] Han J, Kamber M, Pei, J. Data Mining Concepts and Techniques (pp. 15–21). 3rd edn. Morghan Kaupmann, Elsevier: San Francisco, California, United States of America. 2011.
[21] Luu K, Bazin E, Blum MGB. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources. 2017; 17: 67–77.
[22] Whitlock MC, Lotterhos KE. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F(ST). The American Naturalist. 2015; 186: S24–S36.
[23] Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society B: Biological Sciences. 1996; 263: 1619–1626.
[24] Ahrens CW, Rymer PD, Stow A, Bragg J, Dillon S, Umbers KDL, *et al*. The search for loci under selection: trends, biases and progress. Molecular Ecology. 2018; 27: 1342–1356.
[25] Beaumont MA. Adaptation and speciation: what can F(st) tell us? Trends in Ecology & Evolution. 2005; 20: 435–440.
[26] Flanagan SP, Jones AG. Constraints on the FST-Heterozygosity Outlier Approach. The Journal of Heredity. 2017; 108: 561–573.
[27] Pariset L, Joost S, Marsan PA, Valentini A, Econogene Consortium (EC). Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. BMC Genetics. 2009; 10: 7.
[28] Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Molecular Ecology. 2014; 23: 2178–2192.
[29] Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100: 9440–9445.
[30] Viegas MP, Silva CLSP, Moreira JP, Cardin LT, Azevedo VCR, Ciampi AY, *et al*. Genetic diversity and effective population size of two populations of Myracrodruon urundeuva Fr. All., under ex situ conservation. Tree Journal. 2011; 35: 769–779. (In Portuguese)
[31] Nei M. Molecular Evolutionary Genetics. Columbia University Press: New York. 1997.
[32] Cockerham CC, Weir BS. Variance of actual inbreeding. Theoretical Population Biology. 1983; 23: 85–109.
[33] Wright S. The genetical structure of populations. Annals of Eugenics. 1951; 15: 323–354.
[34] Wright S. The interpretation of population structure by F-statistics with special regard to system of mating. Evolution. 1965; 19: 395–420.
[35] Hongyu K, Sandanielo VLM, Oliveira Junior GJ. Principal Component Analysis: theory, interpretations and applications. Engineering and Science. 2015; 5: 83–90. (In Portuguese)
[36] Ferreira MB, Guidolin DGF, Vargas Junior FM, Reis FA, Costa JAA, Vieira ACRD. Pantaneiro genetic group. In Rego, FCA. (ed.) Characterization of carcass and meat from meat lambs. UNOPAR (pp. 163–173). 2nd edn. Editora Científica: Londrina. 2020. (In Portuguese)
[37] Oliveira MF, Ojeda Filho SCF, Hernández I, Leite L, Souza JC, Abreu UGP, *et al*. Phenotypic evaluation of Texel sheep raised in the upper Pantanal. Embrapa Cerrados. 2007. (In Portuguese)
[38] Pérez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J. Microsatellite variation and the differentiation of modern humans. Human Genetics. 1997; 99: 1–7.
[39] Woo J, Suh W, Sung JH. Hair Growth Regulation by Fibroblast Growth Factor 12 (FGF12). International Journal of Molecular Sciences. 2022; 23: 9467.
[40] Ochi K, Derfoul A, Tuan RS. A predominantly articular cartilage-associated gene, SCRG1, is induced by glucocorticoid and stimulates chondrogenesis in vitro. Osteoarthritis and Cartilage. 2006; 14: 30–38.
[41] Loftus SK, Lundh L, Watkins-Chow DE, Baxter LL, Pairo-Castineira E, Nisc Comparative Sequencing Program, *et al*. A custom capture sequence approach for oculocutaneous albinism identifies structural variant alleles at the OCA2 locus. Human Mutation. 2021; 42: 1239–1253.
[42] Nefedova VV, Kopylova GV, Shchepkin DV, Kochurova AM, Kechko OI, Borzova VA, *et al*. Impact of Troponin in Cardiomyopathy Development Caused by Mutations in Tropomyosin. International Journal of Molecular Sciences. 2022; 23: 15723.
[43] Crispim BA, Grisolia AB, Seno LO, Egito AA, Vargas Junior FM, Souza MR. Genetic diversity of locally adapted sheep from Pantanal region of Mato Grosso do Sul. Genetics and Molecular Research: GMR. 2013; 12: 5458–5466.
[44] McManus C, Paiva SR, Caetano AR, Hermuche P, Guimarães RF, Carvalho Junior OA, *et al*. Landscape genetics of sheep in Brazil using SNP markers. Small Ruminant Research. 2020; 192: 106239.
[45] Oliveira DP, Oliveira CAL, Martins EN, Vargas Junior FM, Seno LO, Pinto GS, *et al*. Genetic parameters of performance traits in Sul-Mato-Grossenses naturalized sheep. Semina: Ciências Agrárias. 2014; 35: 963–972. (In portuguese)
[46] Garza Hernandez D, Mucha S, Banos G, Kaseja K, Moore K, Lambe N, *et al*. Analysis of single nucleotide polymorphisms variation associated with important economic and computed tomography measured traits in Texel sheep. Animal: an International Journal of Animal Bioscience. 2018; 12: 915–922.
[47] Kumar H, Panigrahi M, Rajawat D, Panwar A, Nayak SS, Kaisa K, *et al*. Selection of breed-specific SNPs in three Indian sheep breeds using ovine 50 K array. Small Ruminant Research. 2021; 205: 106545.