J. Sylhet Agril. Univ. 12(1): 01-15, 2025 *https://doi.org/10.3329/jsau.v12i1.85889*

ISSN: 2308-1597 (Print) 3105-4080 (Online)

Journal of the Sylhet Agricultural University

Journal home page: http://www.jsau.sau.ac.bd

Research Article

A STUDY ON THE GENETIC DIVERSITY OF SEVEN CATTLE (*BOS* SPP.) BREED POPULATION

Darimi Hasin[1]* and Mohammad Zahangir Alam[1]

[1]Dept. of Genetic Engineering and Biotechnology
Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh

Article info

Article history
Received : 06.12.2024
Accepted : 18.05.2025
Published : 30.06.2025

Keywords
Genetic diversity, High-density SNP chips, Admixture analysis, PCA, Inbreeding depression

Corresponding author
Darimi Hasin
E-mail: darimihasinreem@gmail.com

Abstract

Genetic characteristics are the key to carrying out any modification on an animal to achieve desired goals. High-density Bovine SNP chips were used for three hundred and thirteen individuals of seven cattle breeds (Jeju Black Cattle, Brahman, Nelore, Hereford, Holstein, Hanwoo, and Angus) from secondary sources to study their genetic structure and diversity. The study was carried out from June to September 2021. Different parameters, viz. principal component analysis, minor allele frequency, F_{IS}, F_{ST} analysis, admixture analysis, and linkage disequilibrium, were set as the basis of this investigation. Two different versions of PLINK were used to estimate and calculate the diversity parameters. The minor allele frequency in Holstein (0.01) showed the highest level of genetic diversity, and the lowest was found in Brahman cattle. Heterozygosity was the lowest in Brahman (-0.09), and the highest was in Hanwoo (0.42). Hanwoo and Jeju Black Cattle also showed a significant level of linkage disequilibrium (mean LD was 0.63% and 0.71%, respectively). The F_{ST} values showed the most distance between Hereford and Nelore (0.5). The distance between Hanwoo and Holstein was very low (0.07), and the least distance was found between Nelore and Brahman (0.06). Holstein, Hanwoo, and Jeju Black cattle showed minimal admixture in their ancestry. Study revealed a satisfactory genetic history of the seven cattle breeds, which should be insightful for future conservation programs.

Introduction

The complete set of genes, also known as the genome, provides all the functional abilities that a living being needs. A breed population would have a variety of genes that are responsible for the differences in appearance, behavior, meat or milk production and quality, reproductivity, survivability and overall inheritance among its members (Ellegren and Galtier, 2016). The observation of significant breed differences, particularly in farm animals, are applied in selective breeding programs, reinforcing this global concentration on a few breeds. Such initiatives will only pay off if the genetic advancement in breeding programs can be widely communicated (Oldenbroek and Waaij, 2014). It is essential to acquire proper knowledge about the genetic diversity and interrelations of farm breeds to make genetic improvements and understand their environmental adaptation, along with their utilization and conservation (Makina *et al*., 2015).

Cite This Articale
Hasin D and Alam MZ. 2025. A Study on the Genetic Diversity of Seven Cattle (*Bos* spp.) Breed Population. J. Sylhet Agril. Univ. 12(1): 01-15, 2025. https://doi.org/10.3329/jsau.v12i1.85889

Genetic variations can commonly be found when alleles are present with different distribution frequencies in a particular population. Natural evolutionary changes and artificial selective breeding are enabled through genetic variation (Swingland, 2013). Polymorphisms at the DNA level are revealed by various molecular markers, which are very important in the field of animal genetics. Parameters such as pairwise Fst, heterozygosity test, inbreeding coefficient, and admixture analysis of these markers can draw conclusions about the interrelation and genetic variation among animals, providing scope for targeted breeding technologies. There are three most used variation types at the DNA level markers: single nucleotide polymorphisms (SNPs), insertions or deletions (Indels) of various lengths ranging from one to several hundred base pairs, and variable number of tandem repeats (VNTR) (Vignal *et al*., 2002).

SNP chips are now the most frequently used markers. It is possible to find one C instead of a T in every thousand bases of a DNA strand. These small changes are encoded as SNPs (Aitken *et al*., 2004). High-density SNPs play a vast role in studying and improving the genetic characteristics of many animals for economic use and research purposes. The higher the density of these SNPs is, the better the analysis would be. Other markers include microsatellites, allozymes, RAPDs, and their derivatives etc. A variation in electrophoretic mobility is the most typical approach to detecting changes in these properties (Schlötterer, 2004). However, in the case of indigenous breeds or individuals among the same species, e.g., cattle breeds, the majority of the allele frequencies are common. To filter these SNPs, the focus needs to be on the minor allele frequencies. That is why SNPs need to undergo genotyping and quality control (QC) procedures for better statistical analysis (Mustafa *et al*., 2017). Data analysis tools for SNP genotyping and copy number variation (CNV) can examine results for millions of markers and probes to detect sample outliers. They can also give insights into the functional consequences of any genetic variation. QC techniques vary based on the approach employed. These include eliminating loci with a low sequencing depth and loci with a low PHRED-like quality score (*Q*) for NGS genotyping (Pavan *et al*. 2020). To eliminate bias and false signals in genotype-trait association testing, the QC methods must be augmented. These include filtering processes that are either universal in all GWAS experiments or particular to the GWAS population type.

The study was designed to analyze the genetic structure and variability of seven cattle breeds using RStudio and PLINK software on a high-density SNP dataset.

Materials & Methods

The study was conducted in the Department of Genetic Engineering and Biotechnology at Shahjalal University of Science and Technology using high-density SNP data and computer-based software - RStudio and PLINK.

Data Source

Secondary SNP chip data have been used from online sources to study the genetic diversity of a particular population (Alam *et al*., 2021). A total of 45526 SNP of 313 cattle were taken for genetic diversity analysis. The seven breeds included Korean Hanwoo (HW) and Jeju Black cattle (JJBC); Hereford (HF), Angus (AG) cattle; Nelore (NL) and Brahman (BM) cattle that are originated from India, and Holstein (HST) cattle in the population.

Quality Control of Data

The quality control (QC) of single-nucleotide polymorphism (SNPs) filtering was an essential stage in our analysis to avoid possible misleading findings (Pongpanich *et al*., 2010). It helped in assisting in the identification of poor samples that have slipped past raw data and alignment quality control checks, and reduced the rate of false-positive SNP calls. SNP quality control is frequently performed to remove SNPs with insufficient genotyping quality, e.g., Hardy–Weinberg equilibrium, minor allele frequency (MAF), etc., using expert-guided filters based on QC variables (Pongpanich *et al*., 2010). PLINK 1.9 and PLINK 2.0 were

used for the current data analysis. Both versions of the application were downloaded from https://www.cog-genomics.org/plink2 for Windows 64-bit. Another software, RStudio, was used to execute visual figures and plots from the analysis of the dataset.

At first, the directory was set to the folder "plink" on "Command Prompt" where all the data files and the application were kept. Then, it was checked if the application file was intact or not by running a simple command. Typing the 'plink' command and then specifying a file with no further options took a single parameter and looked for raw data files. This command needed two types of files to read: a ped file and a map file. PLINK is set for human data by default. As cattle data was used for this study, it was also specified that by typing --cow. The data file "7_breed_geno" was taken as the input file.

plink --file 7_breed_geno --cow

PED and MAP files were both simple text files that contain genotype information (PED file); one individual per row, and information on the name and position of markers in the PED file (MAP file). The command stated above generated a log file.

The PED files and MAP files were text files. To avoid the hassle of large text files, the raw data file was converted into a binary file using the following command. We used --out to specify the output file that is "7breed_binary".plink --file 7_breed_geno --cow --nonfounders --make-bed --out 7breed_binary

The risk of unfunctional probes and contaminated data was minimized by excluding data where more than 20% of genotype calls were missing for variants. Additionally, we removed any data with more than 20% of genotype calls missing for any individual. Major allele frequency is mostly the same in a population. So, it was focused on the minor allele frequency to limit the variable sites with a minor allele frequency of less than 20 %.plink --bfile 7breed_binary --cow --nonfounders --geno 0.2 --mind 0.2 --maf 0.2 --make-bed --out 7breed_filtered1plink --bfile 7breed_filtered1 --cow --nonfounders --missing --out 7breed_filtered2 --make-bed

The Hardy-Weinberg equilibrium (HWE) test is crucial when there are significantly more or fewer heterozygous calls than predicted. These irregularities could be the result of a systemic variant calling error (Chang, 2020). We set a p-value of 0.00005 in this case. plink --bfile 7breed_filtered2 --cow --nonfounders --hwe 0.00005 --make-bed --out 7breed_hwe

Genetic Diversity Analysis

Multiple parameters were used for the diversity analysis. Principal Component Analysis (PCA), Pairwise F_{ST} Analysis, Admixture analysis, Linkage Disequilibrium (LD), and Heterozygosity test were performed for this study.

Principal Component Analysis (PCA)

A multivariate approach is used for the principal component analysis (PCA) to identify breed associations directly related to allele frequencies. The data from many alleles and loci are summarized into several synthetic variables called principal components (PC) (Edea *et al.*, 2013). PCA was performed to infer relationships between the seven cattle populations using PLINK 2. PCA plots were formed using RStudio 4.0.

plink2 --bfile 7breed_hwe --cow --out pca_7breed –pca

Heterozygosity and Inbreeding Coefficient Analysis

The heterozygosity test is a powerful tool to find out the history and genetic structure of a population, as well as to investigate the level of inbreeding among any population. It is often tested based on the Hardy-Weinberg Equilibrium (Meyermans *et al.*, 2020). Allele frequency was calculated as double the gene frequency, as every genotype consists of two alleles.

This study estimated a comparison between the observed heterozygosity and the expected heterozygosity from each cattle population under the Hardy-Weinberg Equilibrium. Plink v.2 was used to run the commands for heterozygosity tests. The inbreeding coefficient (F_{IS}) was also measured using the heterozygosity values. The equation followed is,

$$F_{IS} = (H_E - H_O)/H_E$$

Here, H_E represents the heterozygosity value that is expected due to Hardy-Weinberg Equilibrium, and H_O represents the value of heterozygosity that is observed from the SNP data. The higher the observed heterozygosity gets, the lower the F_{IS} value would be, which means less inbreeding in the population (Zhivotovsky, 2015).

Genetic Distance and Pairwise F_{ST} Analysis

SNP data-based Pairwise F_{ST} Analysis was done using the SNPRelate package from R (Zheng *et al*., 2012). The algorithms in this package were coded using C/C++. The script for the F_{ST} analysis was collected from GitHub.

Population Structure Analysis

Admixture analysis is a good tool for population structure and ancestry analysis. In this study, 313 individuals of 7 breeds of cattle underwent the admixture analysis process. An R package, LEA (Landscape and Ecological Association), was used for the analysis. To estimate individual ancestry coefficients and ancestral allele frequency, 'snmf' was used (Frichot *et al*., 2014) 'snmf' provided structure-like outputs.

Linkage Disequilibrium or Association Mapping

With the use of linkage disequilibrium, the correlation coefficient (r^2) between two loci was used to estimate the number of recombination events between linked SNPs in each population (LD) (Bhuiyan *et al.*, 2021). PLINK v.1.9 was used for LD with the default settings.

plink --bfile AG –cow --out AG_ld --r2

plink --bfile BM –cow --out BM_ld --r2

plink --bfile HF –cow --out HF_ld --r2

plink --bfile HST –cow --out HST_ld --r2

plink --bfile HW –cow --out HW_ld --r2

plink --bfile JJBC –cow --out JJBC_ld --r2

plink --bfile NL –cow --out NL_ld --r2

In the commands stated above, AG was for Angus, BM was for Brahman, HF was for Hereford, HST was for Holstein, HW was for Hanwoo, JJBC was for Jeju Black Cattle, and NL was for Nelore.

Results

Quality Control

The quality control process removed 62.47% of the variants, leaving only 37.53% that passed the quality filters. This filtered dataset contained 17,088 high-quality variants, which formed the basis for all subsequent analyses (Table 1).

Table 1. An overview of the quality control of the SNP data

	Without Filter Initial Data	With mind≥ 0.2	With geno≥ 0.2	With maf≥0.2	With hwe≥1e-25
Number of SNPs	45526	45526	(45526-685) = 44841	(44841-18931) = 25910	(25910-8822) = 17088
Number of Individuals	313	313	313	313	313
Total Percentage of Missing Data	0%	0%	1.5%	43.09%	62.47%

The number of remaining SNPs and individuals are displayed with their percentages of the data that is missing without and with filters (Table1), where mind is missing individuals; geno, missing genotype; maf, minor allele frequency; hwe, Hardy-Weinberg equilibrium.

Principal Component Analysis (PCA)

To understand the population structure, Principal Component Analysis (PCA) was conducted using Plink v.2, which provided the top ten principal component values by default (Table 2). Among these, the first two principal components explained a significant portion of the genetic variation, with PC1 and PC2 accounting for 34.64% and 28.25%, respectively. R plots of the PCA results showed distinct clustering patterns among the seven cattle breeds. Notably, Angus (AG), Hereford (HF), and Holstein (HST) were closely clustered together, indicating a degree of genetic similarity between these breeds. In contrast, Hanwoo (HW) and Jeju Black Cattle (JJBC) formed their own distinct clusters. Brahman (BM) and Nelore (NL) also clustered closely with each other, but this cluster had a wider distance from others, confirming their genetic dissimilarities (Figure 1).

Table 2. Top Ten Eigen-values of Principal Component Analysis (PCA)

Principal Component Analysis	Eigen-value
PC1	34.6392
PC2	28.2452
PC3	14.9272
PC4	12.0905
PC5	8.43745
PC6	4.79741
PC7	4.10087
PC8	4.01854
PC9	3.19449
PC10	2.70649

This pattern was consistent across several PCA plots, reinforcing the genetic closeness of AG, HF, and HST, as well as the separation of HW and JJBC. BM and NL consistently stood apart, always forming a tight cluster together, highlighting their distinct genetic background. HST showed a slight tendency toward scattering, but it still maintained proximity to AG and HF in the clustering patterns (Figure 1).

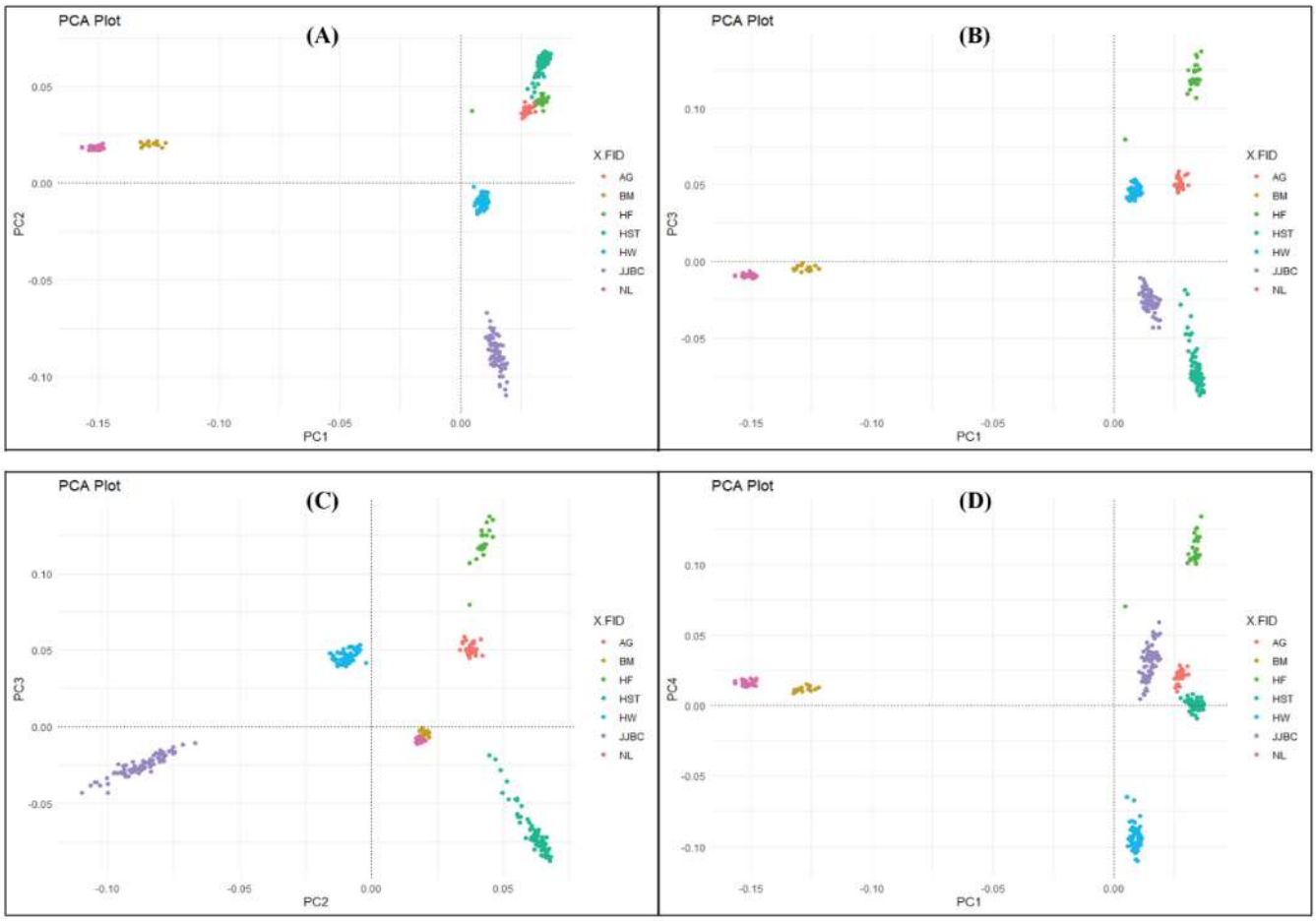


Figure 1. Principal Component Analysis (PCA) for seven cattle breeds. The first and second principal component**(A)**, first and third principal component **(B)**, second and third principal component **(C)**, and first and fourth principal component **(D)** analysis resulted from 17088 SNPs. Each dot corresponds to one individual (AG, Angus; BM, Brahman; HF, Hereford; HST, Holstein; HW, Hanwoo; JJBC, Jeju Black cattle; and NL, Nelore), and each color corresponds to a particular population. The labeling is on the right of each plot. Each cluster represents their genetic similarities, where the distance among the clusters represents the genetic distance among different populations.

Genetic Diversity within Breeds and F-Statistics

The observed heterozygosity (H_O) is higher than the expected heterozygosity (H_E) in every cattle population (Figure 2). The inbreeding coefficient F_{IS} resulted in negative values per population except for NL and BM, which show very high F_{IS} values (Figure 3). A Higher inbreeding coefficient of these two Indian breeds resulted in lower heterozygosity (Figure 2).

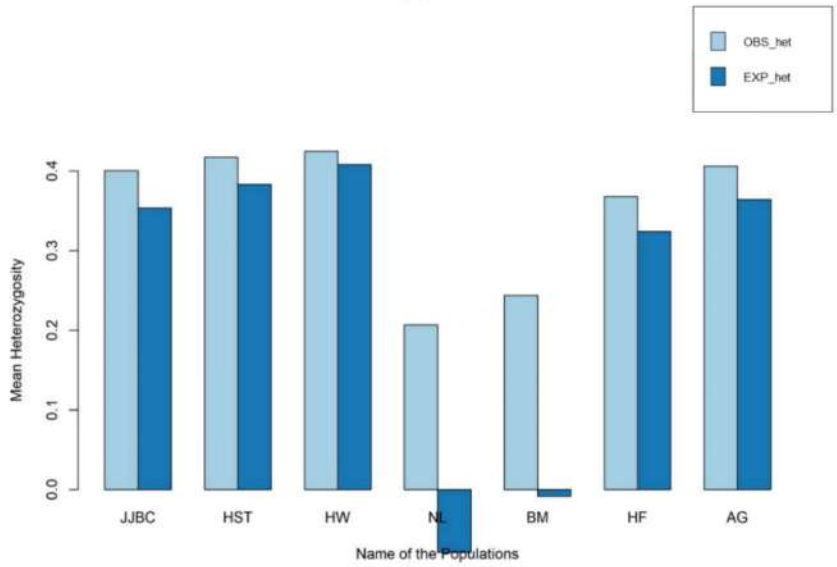


Figure 2. Heterozygosity result of the seven cattle populations. Bar diagram of observed heterozygosity (H_O) vs. expected heterozygosity (H_E) of seven cattle populations. Observed heterozygosity (H_O) is labeled in light blue color, and expected heterozygosity (H_E) is labeled in dark blue color, showing positive results in all the other breeds except NL and BM.

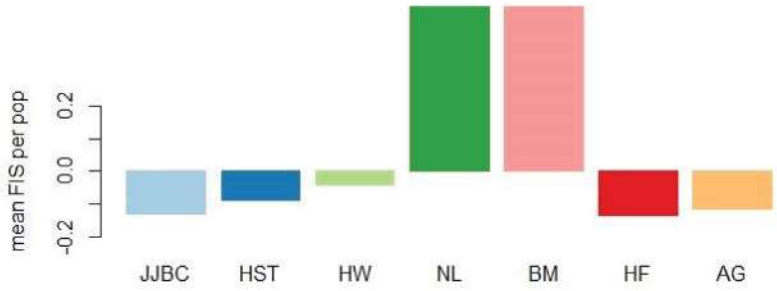


Figure 3. Mean inbreeding co-efficient, F_{IS} for seven cattle breeds. All the breeds other than NL and BM showed a negative inbreeding coefficient.

Genetic Distance and Allele Frequencies

Pairwise F_{ST} values showed the genetic distance among cattle breeds, with the intensity of color as the measure of genetic distance. Darker color represents wider genetic distance, while lighter color means more similarities (Figure 4).

The MAF analysis in this study successfully identified rare and common variants across breeds, revealing genetic diversity and potential selection signatures. The varying allele distributions among breeds indicate differences in genetic makeup, supporting the study's objective of understanding breed-specific genetic variation. NL and BM exhibited the highest allele expressions at low frequencies (MAF = 0.0–0.1), indicating a prevalence of rare alleles. In contrast, AG, HF, HST, HW, and JJBC showed increasing genetic diversity with higher allele frequencies (MAF = 0.4 – 0.5). Among them, HST displayed the highest overall diversity (Figure 5).

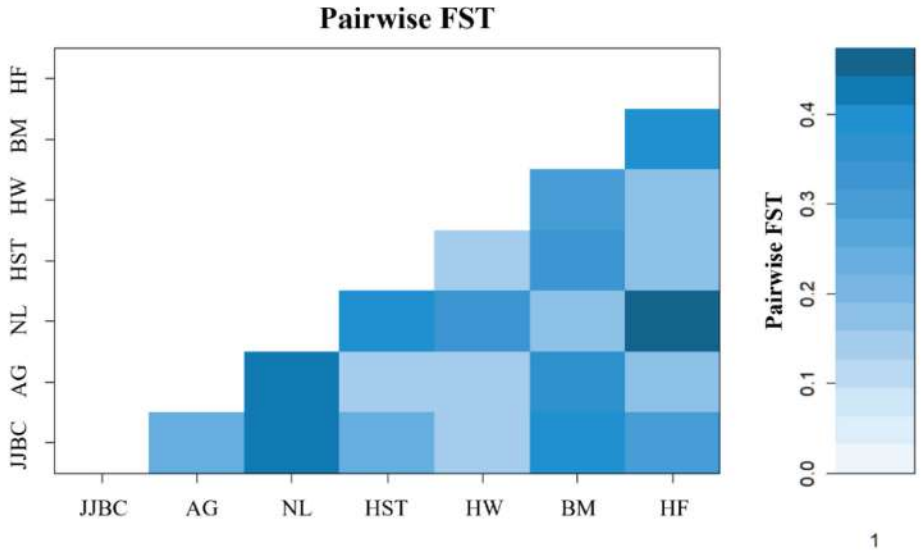


Figure 4. Pairwise F_{ST} values for seven cattle breeds. The intensity of the color ranging from 0.0 to 0.5 represents the F_{ST} values or genetic distance among the breeds. White indicates $F_{ST} = 0.0$, and the darkest blue indicates $F_{ST} = 0.5$. NL-HF has the most genetic distance, whereas breeds like HW-AG and NL-BM have the least distance.

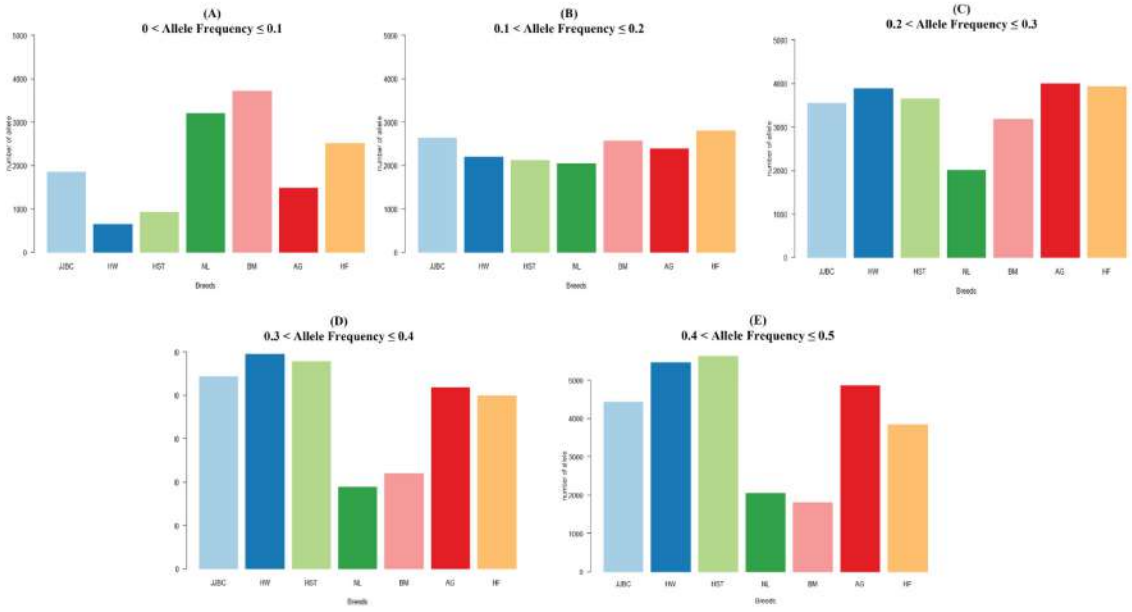


Figure 5. Minor allele frequency distribution (MAF) across the seven cattle breeds. The bar plots represent the number of alleles within specific frequency ranges for various breeds: JJBC, HW, HST, NL, BM, AG, and HF. The panels (A–E) depict different allele frequency intervals: (A) 0 < Allele Frequency ≤ 0.1, (B) 0.1 < Allele Frequency ≤ 0.2, (C) 0.2 < Allele Frequency ≤ 0.3, (D) 0.3 < Allele Frequency ≤ 0.4 and (E) 0.4 < Allele Frequency ≤ 0.5. Each color represents a different breed, and the height of the bars indicates the number of alleles within the given frequency range for each breed.

Population Structure Analysis

The population structure of the seven cattle breeds was analyzed to determine the optimal number of genetic clusters using cross-validation (CV). The cross-validation error decreased with an increasing number of ancestral populations, indicating a better fit of the population model at higher K values (Figure 6).

The admixture analysis uncovered genetic contributions from multiple ancestral populations, highlighting breed intermixing and historical gene flow. The varying admixture proportions across breeds provide insights into population structure and genetic diversity. JJBC formed several sub-clusters with minimal admixture from other populations, highlighting its unique genetic background. HW and HST also showed low levels of admixture, indicating relatively pure ancestry. In contrast, BM and NL clustered closely together across different K values. AG and HF showed some degree of shared ancestry but with limited admixture between them. The population sizes inferred from the admixture analysis included 27 individuals for AG, 15 for BM, 25 for HF, 76 for HST, 66 for HW, 78 for JJBC, and 26 for NL (Figure 7). These results provide a clear view of the genetic structure and historical relationships among the studied breeds.

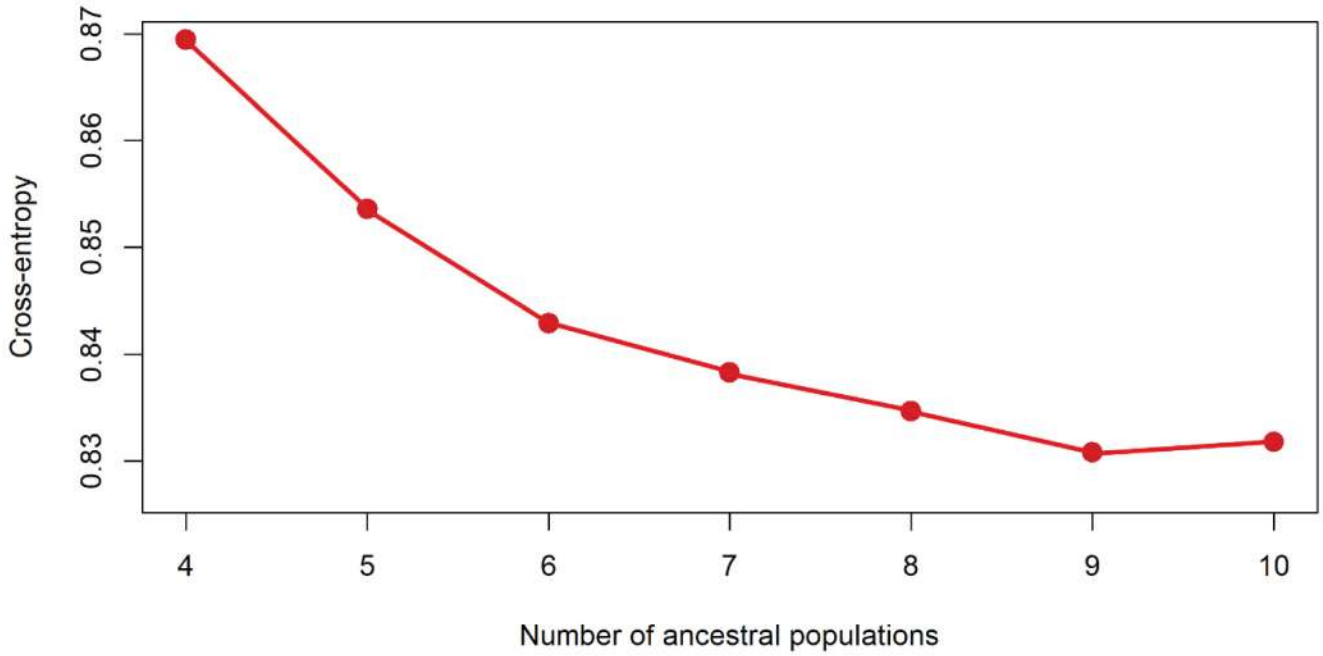
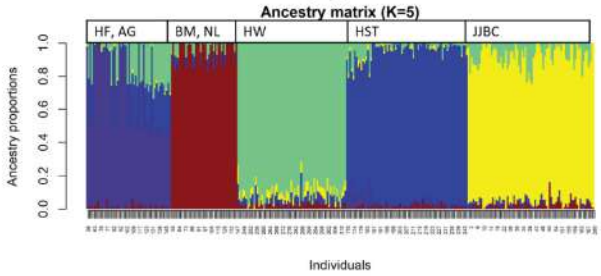
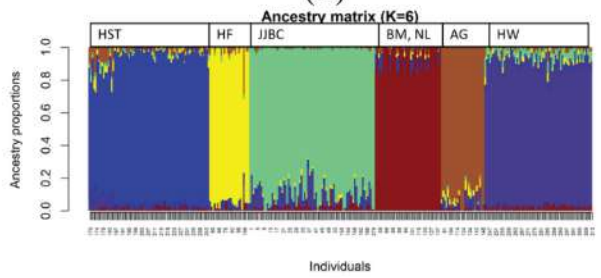


Figure 6. Cross-validation plot for seven cattle breeds.

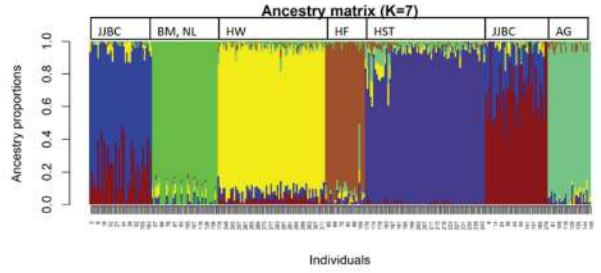
(A)


Ancestry matrix (K=5)
HF, AG
BM, NL
HW
HST
JJBC
Ancestry proportions
0.0 0.2 0.4 0.6 0.8 1.0
Individuals

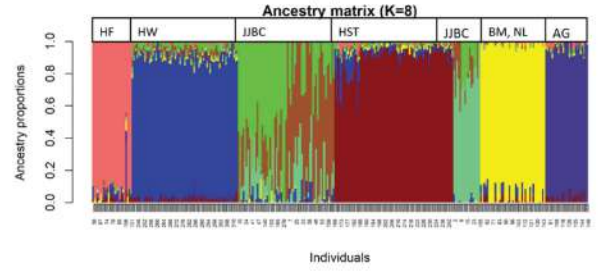
(B)


Ancestry matrix (K=6)
HST
HF
JJBC
BM, NL
AG
HW
Ancestry proportions
0.0 0.2 0.4 0.6 0.8 1.0
Individuals

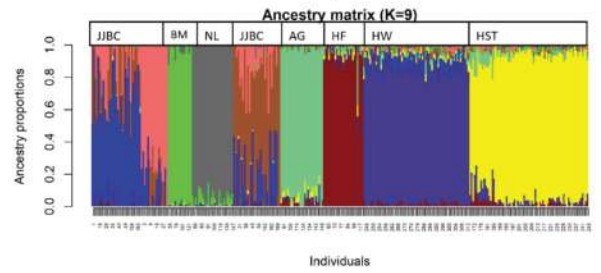
(C)


Ancestry matrix (K=7)
JJBC
BM, NL
HW
HF
HST
JJBC
AG
Ancestry proportions
0.0 0.2 0.4 0.6 0.8 1.0
Individuals

(D)


Ancestry matrix (K=8)
HF
HW
JJBC
HST
JJBC
BM, NL
AG
Ancestry proportions
0.0 0.2 0.4 0.6 0.8 1.0
Individuals

(E)


Ancestry matrix (K=9)
JJBC
BM
NL
JJBC
AG
HF
HW
HST
Ancestry proportions
0.0 0.2 0.4 0.6 0.8 1.0
Individuals

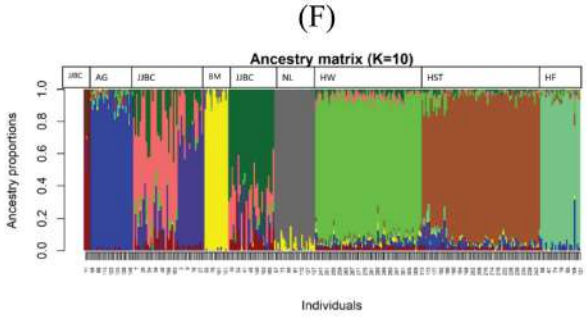


Figure 7. Ancestral clustering of individual cattle high-density SNP data for 313 individuals of seven breeds. Results are shown for estimated ancestral populations with **K = 5, (A); 6, (B); 7, (C); 8, (D); 9, (E); 10, (F).**

Linkage Disequilibrium or Association Mapping

Linkage disequilibrium (LD) analysis revealed genetic linkage across the genome, highlighting selection regions and evolutionary events. Association mapping identified loci linked to traits, suggesting candidate genes and key markers influencing breed-specific characteristics or adaptive traits. Figure 8 showed a rapid decline in LD over shorter genomic distances, while no significant patterns were observed over longer distances.

The pairwise correlation coefficient (r^2) across all populations ranged from 0.2 to 0.7, with LD decreasing as the physical distance between SNP pairs increased. This pattern is consistent with the expected genomic architecture in populations with diverse evolutionary backgrounds and varying degrees of historical recombination. The observed LD decay provides important insights into the genomic resolution and potential mapping power for association studies within these cattle populations (Figure 8).

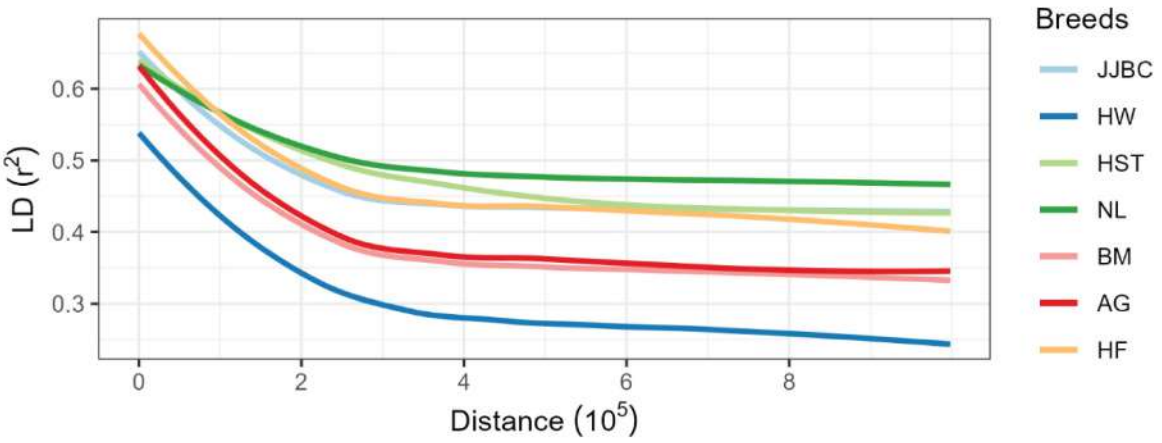


Figure 8. Linkage Disequilibrium (LD) of seven cattle populations. LD decay of the seven cattle breeds was plotted according to the mean pairwise correlation coefficient (r^2) over a genomic distance of 0 to 1000.

Table 3. Mean values of Linkage Disequilibrium (r2) of seven cattle populations

Populations	Mean LD (r^2)
AG	0.4925
BM	0.4722
HF	0.5425
HST	0.5352
HW	0.3925
JJBC	0.5523
NL	0.5525

Discussion

The study results provide valuable insights into the genetic relationships and diversity patterns across seven cattle populations, revealing both expected and novel patterns of genomic variation. The principal component analysis (PCA) revealed distinct genetic clusters, particularly highlighting the similarity between Brahman (BM) and Nelore (NL) and the genetic uniqueness of Hanwoo (HW) and Jeju Black Cattle (JJBC). According to previous studies, the clustering of the two Zebu breeds, BM and NL, reflects their shared indicine ancestry and close genetic relationship due to common origins in South Asia and subsequent breeding histories in Brazil and the United States. Similarly, the distinct clustering of HW and JJBC highlights their genomic divergence, shaped by isolated breeding histories and differing production goals (Alam *et al*., 2021; Decker *et al*., 2014)

Previous genomic studies have confirmed the distinct genetic status and low diversity of JJBC compared to other Korean breeds. JJBC exhibited the lowest heterozygosity and a small effective population size, indicating potential inbreeding and vulnerability, compared to HW. Principal component analysis (PCA) of this study consistently positions JJBC as a distinct cluster, reinforcing its genetic divergence (Lim *et al*., 2014). Sharma *et al*. (2016) also reported lower genetic diversity in JJBC than in Brown and Brindle Hanwoo, indicating greater genetic isolation and potential inbreeding in JJBC compared to other Korean cattle breeds.

Our findings showed that BM and NL had a high proportion of low-frequency alleles (MAF <0.1), consistent with tropical cattle breeds like Sahiwal and Gir. This pattern reflects historical bottlenecks and selective breeding for heat tolerance, parasite resistance, and adaptation to low-quality forage (Ahmad *et al*., 2019). In contrast, Holstein exhibited a higher proportion of intermediate- and high-frequency alleles (MAF 0.3–0.5), aligning with European dairy breeds that have undergone intensive selection for milk production and maintain larger effective population sizes (Makina *et al*., 2015). Hanwoo presented a more complex MAF distribution with fewer low-frequency alleles than previously reported. This may result from differences in sample origin, genotyping platforms, or evolving breeding strategies. Its selection for meat quality traits likely increased the frequency of economically important alleles while maintaining genomic variability. The heterozygosity results showed that JJBC maintained relatively high genetic diversity (HE = 0.36), aligning closely with values reported by (Alam *et al*., 2021). However, earlier work by (Lee *et al*., 2014) estimated lower heterozygosity (HE = 0.29), possibly reflecting a difference in sampling, as Lee's study focused on cattle from more geographically isolated herds in Jeju Island. This discrepancy highlights how sampling strategies, marker density, environmental adaptation, and breeding objectives collectively influence genetic diversity across cattle breeds.

A much lower expected heterozygosity in BM (HE = -0.09) and NL (HE = -0.01) was observed, which is unusual. Consequently, the inbreeding coefficients (FIS) were notably high, reflecting their historical use of closed populations and selective breeding practices in tropical regions. This pattern matches findings from the large-scale indicine genome project, which reported increased FIS in commercial NL herds (Utsunomiya *et al*., 2019).

Similarly, McTavish *et al*. (2013) reported that indicine cattle, particularly those adapted to tropical environments, experienced strong population bottlenecks during domestication and subsequent breeding, further contributing to their elevated inbreeding levels (McTavish *et al*., 2013; McTavish, 2013). Analysis of indicine cattle genomes also revealed reduced genome-wide diversity relative to taurine breeds, partly due to these historical constraints (McTavish *et al*., 2013; Decker *et al*., 2014).

In contrast, Hanwoo and Holstein exhibited the lowest inbreeding levels, a consequence of structured breeding programs and periodic introduction of novel genetic material through artificial insemination (AI) and crossbreeding. The Holstein breed, in particular, benefits from extensive international genetic exchange, helping to maintain diversity and reduce FIS (Decker *et al*., 2014). Decker *et al*. (2014) further highlighted that structured breeding and well-documented pedigrees in taurine breeds have led to more effective management of genetic diversity compared to indicine breeds.

Nelore and Brahman showed unusually negative inbreeding coefficients, suggesting excess heterozygosity despite their histories of closed populations and selection in tropical regions. This contrasts with earlier studies on indicine breeds (e.g., Gir) that reported positive inbreeding values, consistent with controlled breeding and reduced diversity, and with reports of strong bottlenecks in tropically adapted indicine cattle (Srivastava *et al*., 2019; Ahmad *et al*., 2019; O'Brien *et al*., 2015; E. J. B. McTavish, 2013). We suspect that the surplus of heterozygosity may be due to bias in the SNP array, since these arrays are usually designed for taurine cattle; they may not accurately represent indicine genomes (Makina *et al*., 2015). On the other hand, Hanwoo and Holstein showed positive inbreeding coefficients, reflecting selection for marbling and milk yield. Hanwoo's moderate heterozygosity shows its shift from draft to beef use. These findings illustrate how selection history shapes diversity and emphasize the need for breed-specific genotyping tools for indicine cattle to avoid ascertainment bias and misinterpretation (Utsunomiya *et al*., 2019; de las Heras-Saldana *et al*., 2019; Choi *et al*., 2014).

The admixture analysis further clarified these population histories, showing that even at K = 9, BM and NL remained largely within a shared cluster, while Hanwoo, Jeju Black, and Holstein each maintained distinct genetic identities. Notably, Jeju Black exhibited greater internal heterogeneity, splitting into multiple components at higher K, consistent with Lim *et al*. (2020), who suggested that modern Jeju Black populations have incorporated genes from mainland Korean cattle (Lim *et al*., 2020). Linkage disequilibrium (LD) was highest in Nelore and Jeju Black (r^2 = 0.5525 and 0.5523), and lowest in Hanwoo (r^2 = 0.3925), reflecting the inverse relationship between LD and genetic diversity typical of breeds shaped by bottlenecks or founder events (Gurgul *et al*., 2016). Hanwoo's lower LD reflects its higher genetic diversity, consistent with its diverse breeding history, which spans both indigenous and imported bloodlines (Lee *et al*., 2014).

Conclusion

This study provided an overview of the genetic relationship among seven cattle breeds from different landraces. The indicine cattle breeds, i.e., Brahman and Nelore, were genetically strongly related and showed the lowest level of genetic diversity. The study also revealed the strong relationship between Hanwoo and Jeju Black cattle. The Korean native cattle breeds showed very high genetic diversity compared to the others. Holstein showed a completely different ancestry in the admixture analysis, while Angus and Hereford were relatively closer to each other. The findings reported the basic genomic characteristics of the seven cattle breeds that will help conserve the information for future genetic modification programs.

Acknowledgements

The author would like to express her gratitude to Fardin Zahid, a fellow from the Department of Genetic Engineering and Biotechnology (GEB), Shahjalal University of Science and Technology (SUST), who has helped greatly throughout the entire duration of this research project.

References

Ahmad, SF, Kour G, Singh A, & Gulzar M. 2019. *Animal Genetic Resources of India—An Overview*. *9*, 1–12. https://doi.org/10.5455/ijlr.20181025013931.

Aitken N, Smith S, Schwarz C, & Morin PA. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: A targeted-gene approach. *Molecular Ecology*, *13*(6), 1423–1431. https://doi.org/10.1111/j.1365-294X.2004.02159.x.

Alam MZ, Lee YM, Son HJ, Hanna LH, Riley DG, Mannen H, Sasazaki S, Park, SP, & Kim JJ. 2021. Genetic characteristics of Korean Jeju Black cattle with high density single nucleotide polymorphisms. *Animal Bioscience*, *34*(5), 789–800. https://doi.org/10.5713/ajas.19.0888.

Oldenbroek K, Waaij LVD. 2014. *Animal breeding and genetics for BSc students: Textbook animal breeding—Wageningen University & Research*. (n.d.). Retrieved August 25, 2025, from https://research.wur.nl/en/publications/animal-breeding-and-genetics-for-bsc-students-textbook-animal-bre.

Choi JW, Liao X, Stothard P, Chung WH, Jeon HJ, Miller SP, Choi SY, Lee JK, Yang B, Lee KT, Han KJ, Kim HC, Jeong D, Oh JD, Kim N, Kim TH, Lee HK, & Lee SJ. 2014. Whole-Genome Analyses of Korean Native and Holstein Cattle Breeds by Massively Parallel Sequencing. *PLOS ONE*, *9*(7), e101127. https://doi.org/10.1371/journal.pone.0101127.

De las Heras-Saldana S, Clark SA, Duijvesteijn N, Gondro C, van der Werf JHJ, & Chen Y. 2019. Combining information from genome-wide association and multi-tissue gene expression studies to elucidate factors underlying genetic variation for residual feed intake in Australian Angus cattle. *BMC Genomics*, *20*(1), 939. https://doi.org/10.1186/s12864-019-6270-4.

Decker JE, McKay SD, Rolf MM, Kim J, Alcalá AM, Sonstegard TS, Hanotte O, Götherström A, Seabury CM, Praharani L, Babar ME, Regitano LC de A, Yildiz MA, Heaton MP, Liu WS, Lei CZ, Reecy JM, Saif-Ur-Rehman M, Schnabel RD, & Taylor JF. 2014. Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLOS Genetics*, *10*(3), e1004254. https://doi.org/10.1371/journal.pgen.1004254.

Frichot E, Mathieu F, Trouillon T, Bouchard G, & François O. 2014. Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, *196*(4), 973–983. https://doi.org/10.1534/genetics.113.160572.

Gurgul A, Szmatoła T, Topolski P, Jasielczuk I, Żukowski K, & Bugno-Poniewierska M. 2016. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. *Journal of Applied Genetics*, *57*(4), 527–530. https://doi.org/10.1007/s13353-016-0337-6.

Lee SH, Park BH, Sharma A, Dang CG, Lee SS, Choi TJ, Choy YH, Kim HC, Jeon KJ, Kim SD, Yeon SH, Park SB, & Kang HS. 2014. Hanwoo cattle: Origin, domestication, breeding strategies and genomic selection. *Journal of Animal Science and Technology*, *56*(1), 2. https://doi.org/10.1186/2055-0391-56-2.

Lim D, Lee SH, Gondro C, Choi BH, Chai HH, Strucken E, Cho YM, Jang GW, Hong SK, & Kim TH. 2014. *Identification of selection footprints in a brown Hanwoo (Korean cattle) population for production traits*. https://rune.une.edu.au/web/handle/1959.11/19860.

Makina SO, Muchadeyi FC, van Marle-Köster E, Taylor JF, Makgahlela ML, & Maiwashe A. 2015. Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genetics Selection Evolution*, *47*(1), 92. https://doi.org/10.1186/s12711-015-0173-x.

McTavish EJB. 2013. *Estimating population histories using single-nucleotide polymorphisms sampled throughout genomes*. http://hdl.handle.net/2152/21925.

McTavish EJ, Decker JE, Schnabel RD, Taylor JF, & Hillis DM. 2013. New World cattle show ancestry from multiple independent domestication events. *Proceedings of the National Academy of Sciences*, *110*(15), E1398–E1406. https://doi.org/10.1073/pnas.1303367110.

Meyermans R, Gorssen W, Buys N, & Janssens S. 2020. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, *21*(1), 94. https://doi.org/10.1186/s12864-020-6463-x.

Mustafa H, Eiusoo K, Heather HJ, Ajmal A, Riley D, Pasha NT, Ali A, Javed K, & Sonstegard TS. 2017. Genome-wide SNPs analysis of indigenous zebu breeds in Pakistan. *Biotechnology in Animal Husbandry*, *33*(1), 13–25.

O'Brien AMP, Höller D, Boison SA, Milanesi M, Bomba L, Utsunomiya YT, Carvalheiro R, Neves HH, da Silva MV, VanTassell CP, Sonstegard TS, Mészáros G, Ajmone-Marsan P, Garcia F, & Sölkner J. 2015. Low levels of taurine introgression in the current Brazilian Nelore and Gir indicine cattle populations. *Genetics Selection Evolution*, *47*(1), 31. https://doi.org/10.1186/s12711-015-0109-5.

Pavan S, Delvento C, Ricciardi L, Lotti C. 2020. Recommendations for Choosing the Genotyping Method and Best Practices for Quality Control in Crop Genome-Wide Association Studies. *Frontiers in Genetics, 11*, Article 447. https://doi.org/10.3389/fgene.2020.00447.

Pongpanich M, Sullivan PF, & Tzeng JY. 2010. A quality control algorithm for filtering SNPs in genome-wide association studies. *Bioinformatics*, *26*(14), 1731–1737. https://doi.org/10.1093/bioinformatics/btq272.

Schlötterer C. 2004. The evolution of molecular markers—Just a matter of fashion? *Nature Reviews Genetics*, *5*(1), 63–69. https://doi.org/10.1038/nrg1249.

Sharma A, Lee SH, Lim D, Chai HH, Choi BH, & Cho Y. 2016. A genome-wide assessment of genetic diversity and population structure of Korean native cattle breeds. *BMC Genetics*, *17*(1), 139. https://doi.org/10.1186/s12863-016-0444-8.

Swingland IR. 2013. Biodiversity, Definition of. In S. A. Levin (Ed.), *Encyclopedia of Biodiversity (Second Edition)* (pp. 399–410). Academic Press. https://doi.org/10.1016/B978-0-12-384719-5.00009-5.

Utsunomiya YT, Milanesi M, Fortes MRS, Porto-Neto LR, Utsunomiya ATH, Silva MVGB, Garcia JF, & Ajmone-Marsan P. 2019. Genomic clues of the evolutionary history of Bos indicus cattle. *Animal Genetics*, *50*(6), 557–568. https://doi.org/10.1111/age.12836.

Vignal A, Milan D, SanCristobal M, & Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, *34*(3), 275. https://doi.org/10.1186/1297-9686-34-3-275.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, & Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics (Oxford, England)*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606.

Zhivotovsky LA. 2015. Relationships Between Wright's FST and FIS Statistics in a Context of Wahlund Effect. *Journal of Heredity*, *106* (3), 306–309. https://doi.org/10.1093/jhered/esv019.