Article

Not peer-reviewed version

Selection Signatures at PRLR and Heat Shock Protein Genes Reveal the Genetic Basis of Heat Stress Adaptation in Cattle

Hamid Mustafa *, Mohammed Al-Abri, Waqas Akram, Haiba Kaul, Waqas Ahmad Khan, Jong-Joo Kim, Muhammad Bilal Bin Majeed

Posted Date: 22 January 2026

doi: 10.20944/preprints202601.1602.v1

Keywords: heat stress adaptation; PRLR; heat shock protein genes; selection signatures; climate resilient cattle

 doi:10.20944/preprints202601.1602.v1

Article

Selection Signatures at PRLR and Heat Shock Protein Genes Reveal the Genetic Basis of Heat Stress Adaptation in Cattle

Hamid Mustafa [1,*], Mohammed Al-Abri [2], Waqas Akram [3], Haiba Kaul [1], Waqas Ahmad Khan [4], Jong-Joo Kim [5] and Muhammad Bilal Bin Majeed [1]

[1] Department of Animal Breeding and Genetics, University of Veterinary and Animal Sciences (UVAS), Lahore–54000, Punjab, Pakistan
[2] Department of Animal and Veterinary Sciences, College of Agricultural and Marine Sciences, Sultan Qaboos University, Muscat 123, Oman
[3] Maxim Agri Private Limited, 7-B, Aziz Avenue, Gulberg 5, Canal Bank Road, Lahore, Pakistan
[4] Department of Biotechnology, University of Sargodha, Sargodha, Pakistan
[5] Department of Biotechnology, Yeungnam University, Gyeongsan, Gyeongbuk 712749, Republic of Korea
* Correspondence: hamidmustafa@uvas.edu.pk; Tel.: +92-308-00-3-99-31

Abstract

Background: Heat stress is a critical constraint to cattle productivity, health, and reproductive efficiency in tropical and subtropical regions, with severity increasing under climate change. Indigenous Bos *indicus* cattle display superior thermotolerance compared with high-producing Bos *taurus* breeds, yet the genomic basis of this adaptation remains incompletely resolved. Genes regulating endocrine signaling and cellular stress responses particularly the prolactin receptor (PRLR) and heat shock protein (HSP) families are strong candidates for heat stress adaptation. **Methods:** High-density SNP genotype data were analyzed from five cattle breeds representing contrasting climatic adaptation: Sahiwal, Cholistani, Brahman, Holstein Friesian, and Jersey. Population structure was examined using principal component analysis and model-based ancestry inference. Genetic differentiation was quantified using pairwise fixation index (F_{ST}) estimates. Genome-wide selection signatures were detected using a sliding-window F_{ST} approach (50 kb windows, 10 kb step size), and candidate regions were functionally annotated to identify genes associated with thermoregulation and cellular stress response. **Results:** Clear genetic separation between Bos *indicus* and Bos *taurus* breeds was observed, with low differentiation among indigenous cattle. Genome-wide scans revealed multiple regions exceeding the 99th and 99.9th percentile thresholds of the F_{ST} distribution, indicating strong selection signals. Prominent signatures were detected at the PRLR locus and within key heat stress–related genes, including HSPB1, HSPA1A, HSF1, ATP2B1, and FGF5. Several highly polymorphic SNPs within PRLR were located in intronic and nested gene regions, suggesting potential regulatory roles influencing endocrine signaling and cellular homeostasis under heat stress. **Conclusion:** This study provides compelling genomic evidence that selection at PRLR and heat shock protein genes forms a central genetic basis for heat stress adaptation in cattle. The identified loci offer robust targets for marker-assisted and genomic selection. Incorporation of these markers into national breeding programs and livestock policies can accelerate the development of climate-resilient cattle populations, strengthen conservation of indigenous genetic resources, and enhance the long-term sustainability of dairy and beef production systems under rising global temperatures.

Keywords: heat stress adaptation; PRLR; heat shock protein genes; selection signatures; climate resilient cattle

 doi:10.20944/preprints202601.1602.v1

Introduction

Climate change poses a substantial challenge to the sustainability of dairy production systems, particularly through rising ambient temperatures and increasing climatic variability [1,2]. Heat stress is a major environmental constraint in tropical and subtropical regions, where it adversely affects feed intake, milk yield, reproductive efficiency, animal health, and welfare [3]. In Pakistan, the dairy sector is a key contributor to agricultural value addition and food security, yet it is increasingly exposed to prolonged heat waves, elevated temperatures, and water scarcity [4]. These stressors underscore the need to better understand the biological mechanisms underlying heat tolerance to support sustainable dairy production in hot environments.

Cattle populations differ markedly in their capacity to cope with thermal stress. Indigenous Bos *indicus* milch-type breeds, including Sahiwal and Cholistani cattle, form the foundation of Pakistan's dairy sector and exhibit superior heat tolerance due to long-term adaptation to hot and arid environments. In contrast, exotic dairy breeds such as Holstein Friesian and Jersey, together with Brahman cattle introduced through international germplasm exchange, originate from temperate or subtropical breeding programs and are commonly used to improve milk and meat production. Under tropical production conditions, these imported breeds generally show greater susceptibility to heat stress and associated production losses [5]. Differences in thermoregulatory efficiency between indicine and taurine cattle are partly attributable to variation in morphological traits, physiological responses, and cellular stress response mechanisms [6,7].

Heat tolerance is a complex, polygenic trait involving multiple biological pathways, including endocrine regulation, cellular stress response, energy metabolism, and immune function [8,9]. At the molecular level, genes encoding heat shock proteins (HSPs) play a central role in maintaining cellular homeostasis during thermal stress by stabilizing proteins and preventing cellular damage [10]. In addition, the prolactin receptor (PRLR) gene has been implicated in thermoregulation through endocrine signaling and is associated with adaptive phenotypes such as the slick hair coat. Collectively, PRLR and key HSP genes (e.g., HSPB1, HSPA1A, and HSF1) represent strong functional candidates underlying genetic adaptation to heat stress in cattle [11].

Advances in high-density single nucleotide polymorphism (SNP) genotyping and population genomics have enabled genome-wide detection of selection signatures, facilitating identification of loci exhibiting elevated genetic differentiation due to natural or artificial selection [12]. Fixation index (F_{ST})-based analyses are widely used to quantify genetic divergence between populations exposed to contrasting environmental conditions [13,14]. Accordingly, the objective of this study was to identify genomic regions and candidate genes associated with heat stress adaptation by detecting selection signatures across indigenous milch-type breeds (Sahiwal and Cholistani) and imported exotic breeds (Holstein Friesian, Jersey, and Brahman) using high-density SNP genotype data. We expected that indigenous Bos *indicus* breeds would cluster closely in population structure analyses and exhibit low pairwise genetic differentiation, whereas comparisons with imported exotic breeds would show pronounced genome-wide F_{ST} divergence, with the strongest selection signatures localized to the PRLR locus and heat shock protein genes.

Materials and Methods

Ethics Statement

This study was conducted using existing genotype data generated under the Higher Education Commission (HEC), Pakistan–funded project (NRPU No. 16844). No new animal handling, sampling, or experimental procedures were performed specifically for this study; therefore, additional ethical approval was not required.

Data

Genotype data from a total of 200 cattle were analysed. The dataset comprised indigenous Bos indicus milch-type breeds, including Sahiwal (n = 36) and Cholistani (n = 34), as well as imported

 doi:10.20944/preprints202601.1602.v1

exotic breeds, namely Holstein Friesian (n = 60), Jersey (n = 40), and Brahman (n = 30). Only genotype data were used in the present analyses.

Genotyping and Quality Control

All animals were genotyped using the Illumina Bovine HD BeadChip (777K SNP array). Genotype data were processed and filtered using PLINK v1.9 [15] following standard quality control procedures to ensure data reliability and consistency across breeds.

At the individual level, samples with a genotype call rate below 95% were excluded from further analyses. At the marker level, SNPs were removed if they exhibited a call rate below 95%, a minor allele frequency (MAF) less than 0.05, or significant deviation from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) within breeds. Only autosomal SNPs passing all QC criteria were retained. The filtered dataset was subsequently used for population structure analyses and genome-wide detection of selection signatures.

Population Structure Analysis

Population genetic structure was assessed using principal component analysis (PCA) and model-based clustering. Prior to structure analysis, SNPs were pruned for linkage disequilibrium (LD) using PLINK v1.9 with a sliding-window approach (--indep-pairwise 50 5 0.2), resulting in 46, 700 SNPs retained for population structure inference. PCA was performed using PLINK v1.9 and visualized in the R statistical environment, with the first two principal components used to examine genetic relationships among breeds.

Model-based clustering was conducted using STRUCTURE v2.3.4 under the admixture ancestry model with correlated allele frequencies. Analyses were performed using the LD-pruned SNP dataset with the number of ancestral populations (K) ranging from 2 to 5. Each run included a burn-in period of 20,000 iterations followed by 100,000 Markov chain Monte Carlo (MCMC) iterations. Additional STRUCTURE analyses were performed at the chromosomal level to investigate genome-wide distribution of indicine and taurine ancestry among breeds.

Genetic Differentiation and Selection Signature Detection

Genetic differentiation among breeds was quantified using the fixation index (F_{ST}) estimated according to the Weir and Cockerham method, as implemented in PLINK v1.9 [15]. Pairwise F_{ST} values were calculated between all breed combinations to characterize genome-wide genetic divergence between indigenous heat-tolerant breeds (Sahiwal and Cholistani) and imported exotic breeds (Holstein Friesian, Jersey, and Brahman).

Selection signatures were detected using a genome-wide sliding-window F_{ST} approach based on autosomal SNPs. Windowed F_{ST} values were calculated using 50-kb windows with a 10-kb step size to smooth local estimates of genetic differentiation across the genome. Windows containing fewer than a minimum number of SNPs were excluded to avoid unstable estimates. Candidate genomic regions under selection were defined as windows exceeding the 99th and 99.9th percentiles of the empirical genome-wide windowed F_{ST} distribution.

Genes located within or immediately flanking candidate windows were annotated using the ARS-UCD1.2 bovine reference genome and corresponding gene annotation resources. Particular emphasis was placed on candidate regions harboring genes with established roles in thermoregulation and cellular stress response, including the prolactin receptor (PRLR) and heat shock protein genes (e.g., HSPB1, HSPA1A, and HSF1).

FST Calculation: Weir and Cockerham Method

FST measures genetic differentiation between populations. The formula by [16] was used to estimate FST for each SNP:

 doi:10.20944/preprints202601.1602.v1

$$FST = \frac{MSP - MSG}{MSP + (r - 1).MSG}$$

Where:

- **MSP** = mean square among populations (between population variance),
- **MSG** = mean square within populations (within population variance),
- **r** = number of populations.

Pairwise FST values were calculated between the heat-tolerant breeds (Sahiwal, Cholistani, Gir) and non-heat-tolerant breeds (Holstein Friesian, Brahman, Jersey) using Variant Cell Format tools. The pairwise FST values provided an estimate of genetic differentiation across all SNP loci. Genome-wide FST estimates were used to identify genomic regions under selection pressure.

Sliding Window Analysis of FST

To detect selection sweeps, a sliding window analysis was performed [17]. The genome was divided into 50 kb windows with a step size of 10 kb to smooth the FST estimates. Regions with consistently high FST values indicated areas of selection.

Visualization of FST Results in R

To visualize genome-wide differentiation, a Manhattan plot was generated in R, with SNP positions along the x-axis and FST values on the y-axis. High peaks indicated potential selection signals.

Interpretation of Results

Regions with high FST values, particularly in heat-tolerant breeds, were flagged as potential targets of natural selection related to heat stress. These regions were further investigated for associated genes and their biological roles in heat tolerance.

Results

In this study, genomic analyses including Principal Component Analysis (PCA), Structure analysis, and Pairwise FST comparison were performed using high-density SNP genotype data among the five cattle breeds: Sahiwal (SW), Jersey (JER), Brahman (BRM), Cholistani (CL), and Holstein Friesian (HOL). The analyses reveal patterns of genetic variation, population structure, and breed differentiation.

Population Structure Analysis

The PCA results (Figure 1) reveal that the first two components, PC1 and PC2, accounted for 34.59% and 7.59%, respectively, of the total variation. These results demonstrated a distinct separation between *Bos indicus* and *Bos taurus* breeds. The first principal component (PC1), with the genetic variation value of 34.59%, effectively marked the difference between the South Asian breeds (Brahman, Cholistani, and Sahiwal) and the European breeds (Jersey and Holstein Friesian). The second principal component (PC2), with a genetic variation value of 7.59%, further differentiated within and between the indicine clusters. The proximity of Sahiwal and Cholistani points suggests a close genetic relationship, likely due to shared geographic origin and adaptation to similar environmental pressures, including high ambient temperatures. Among the *Bos indicus* group, the Brahman breed exhibited a slightly separate cluster, likely due to selective crossbreeding and geographic distribution outside South Asia.

 doi:10.20944/preprints202601.1602.v1

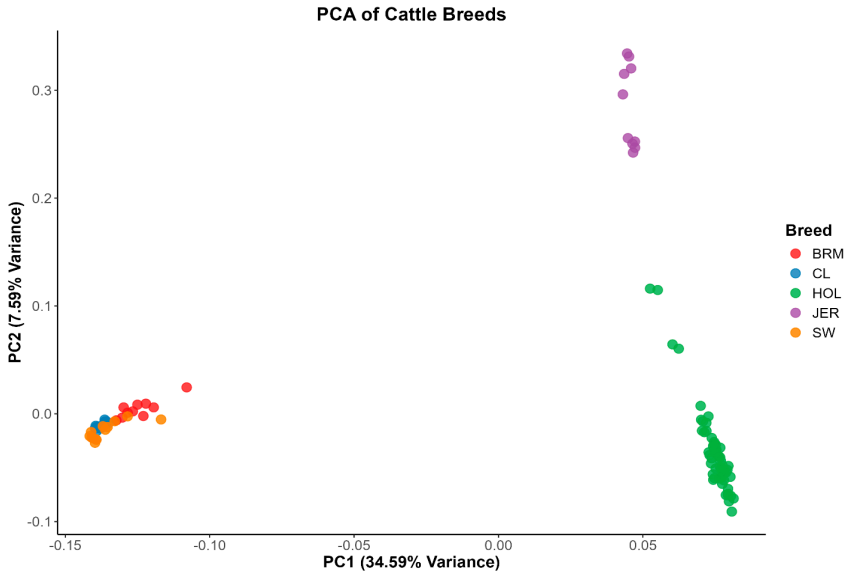

Figure 1. Genetic relationship among the five cattle breeds in this study as inferred by Principal Component Analysis (PCA). Points were colored according to breed: BRM (Brahman), CL (Cholistani), HOL (Holstein Friesian), JER (Jersey), and SW (Sahiwal). The first two components, PC1 and PC2, accounted for 34.59% and 7.59%, respectively, of the total variation. The analysis reveals a clear genetic separation between the European breeds (Holstein Friesian and Jersey) and the breeds of South Asian origin (Brahman, Cholistani, and Sahiwal).

Structure Analysis

The Structure analysis further clarified the population structure. The analysis was performed using varying K values (from K = 2 to K = 5), and the optimal number of genetic clusters was identified at **K=2** whereas the finest informative clustering occurred at **K=5** (Figure 2). The Structure Analysis approach is used to infer the population structure and assess genetic admixture among individuals based on genotype data. As the structure analysis estimates the proportion of the genome originating from reference breeds. At K=2, breeds were vividly differentiated into two distinct genetic clusters corresponding to the two major cattle subspecies: *Bos taurus* (Holstein Friesian and Jersey) and *Bos indicus* (Sahiwal, Cholistani, and Brahman). As the Value of K was increasing, breeds started to separate from each other and show different clusters. At K=3 and K=4, the *Bos taurus* breeds began to separate from each other, while in the *Bos indicus* breeds, the Brahman breed emerged as a distinct cluster within *Bos indicus*. The most informative clustering occurred at k=5, where each breed formed a unique genetic cluster with minimal evidence of admixture. As the value of K increased, the plots reveal continuous subdivision of the genetic variation, revealing more subtle sub-populations or historical gene flow. This genetic distinctiveness supports breed-specific adaptation signatures, which are important for identifying genomic regions linked to heat stress resilience.

 doi:10.20944/preprints202601.1602.v1

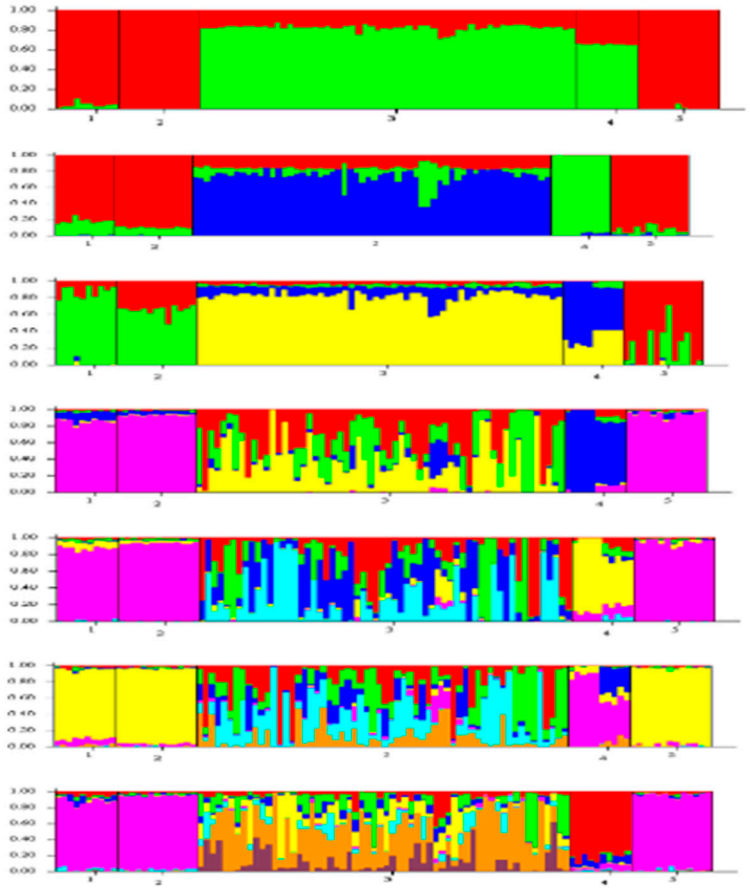

Figure 2. presents a structure analysis of five cattle breeds—Sahiwal, Cholistani, Holstein Friesian, Jersey, and Brahman—revealing their distinct genetic ancestry and relationships across different numbers of assumed ancestral populations (K). At K = 2, the analysis clearly separates the breeds into two major genetic clusters corresponding to their subspecies: Bos taurus (Holstein Friesian and Jersey) and Bos indicus (Sahiwal, Cholistani, and Brahman). As the value of K increases to K = 3 and K = 4, these broader groups begin to show finer genetic differentiation. The Bos taurus breeds remain clearly distinct from one another, while Brahman may form its own cluster, separate from Sahiwal and Cholistani. The most informative split is observed at K = 5, where each of the five breeds likely forms a single, largely unadmixed ancestral cluster. This indicates that all five breeds represent genetically distinct populations. In this representation, each breed's bar plot is predominantly characterized by a unique color. At higher values of K (K = 6 to K = 8), the analysis further subdivides the genetic variation, revealing more subtle subpopulations or traces of historical gene flow.

Fixation Index (FST) Analysis

To detect regions of the genome under positive selection pressure, FST analysis was performed between heat-tolerant indicine breeds (Sahiwal, Brahman, and Cholistani) and non-heat-tolerant taurine breeds (Holstein and Jersey). The fixation index (FST) provides a measure of genetic differentiation between breed pairs. The higher the FST Value, the greater the genetic divergence. In this study, FST values were calculated within 50-kb genomic windows across autosomes, and results were visualized using Manhattan plots for each Pairwise breed comparison. Red and blue horizontal lines indicate the 99.9th and 99th percentile thresholds of the FST distribution, respectively. This allows the identification of genomic regions exhibiting exceptionally high differentiation potential signals of selection or adaptation.

In (Figure 3), the Manhattan Plot shows the genetic differentiation (FST) between the Cholistani and Brahman breeds. Most FST values are below the 99.9th percentile, but peaks on chromosomes 13

 doi:10.20944/preprints202601.1602.v1

and 22 indicate significant divergence despite overall closer genetic relatedness. These regions may reflect adaptation to slightly different environmental pressures or historical breeding practices.

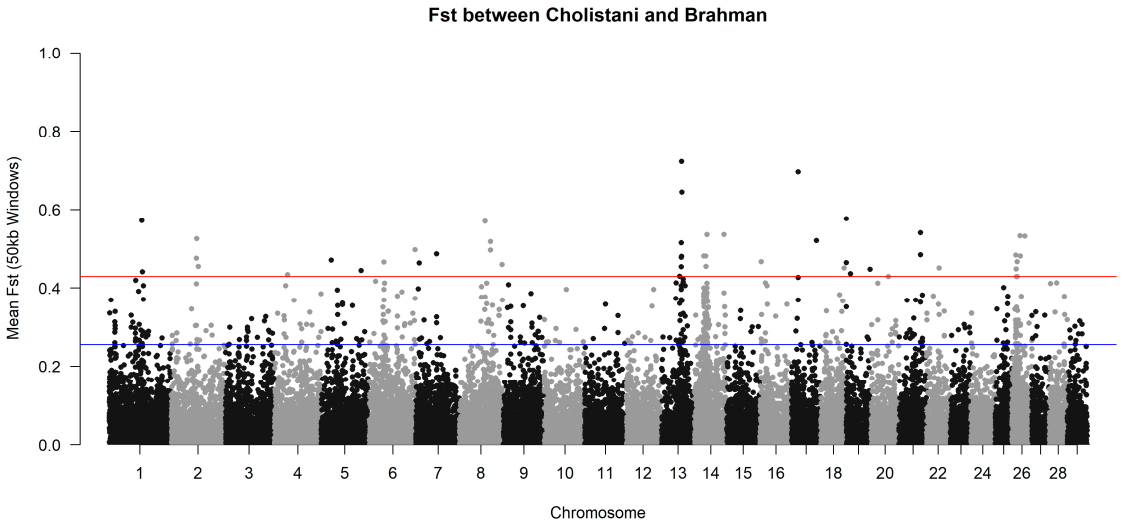

Figure 3. This figure is a Manhattan plot showing the genetic differentiation (FST) between the Cholistani and Brahman cattle breeds.

Manhattan plot (Figure 4) visualizing the FST between Cholistani and Holstein breeds. With many windows exceeding both the 99th and 99.9th percentile thresholds, the plot reveals high-level differentiation. This broad genomic divergence suggests that these two distinct subspecies origins (indica vs taurus) are adapted to contrasting climates, which has driven extensive allele frequency differences.

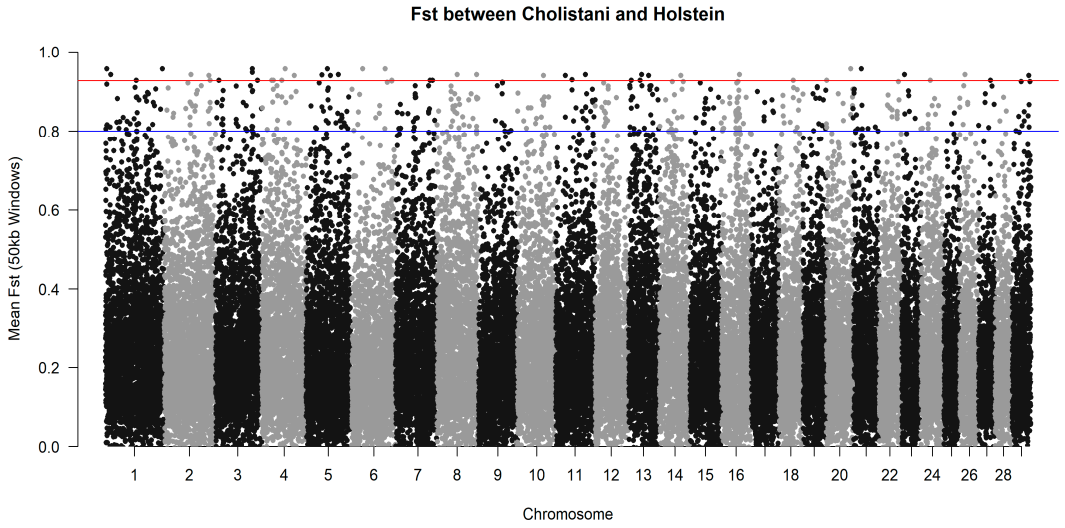

Figure 4. This figure is a Manhattan plot visualizing the genetic differentiation (FST) between the Cholistani and Holstein cattle breeds across the genome.

Manhattan plot (Figure 5) showing FST between Cholistani and Jersey Breeds. High FST values genome-wide, and a large proportion of windows exceeding both thresholds, reflect their distinct

 doi:10.20944/preprints202601.1602.v1

evolutionary histories. Given the jersey's selection focus on milk fat percentage and Cholistani adaptation to arid climates, a clear high differentiation was expected.

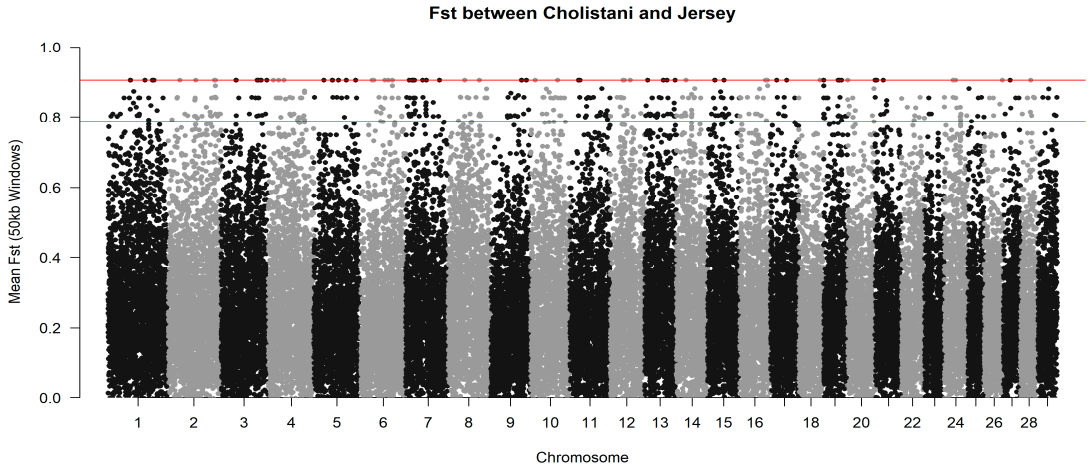

Figure 5. Manhattan plot of genetic differentiation (FST) between the Cholistani and Jersey cattle breeds.

In (Figure 6) Manhattan plot revealing FST between Holstein and Brahman breeds. Extensive differentiation across almost the entire genome and surpassing thresholds. This pattern is expected due to the considerable geographic, environmental, and evolutionary separation between these breeds. The divergence likely represents the combined effects of adaptation to tropical vs temperate climates, selection for different production traits, and limited historical gene flow.

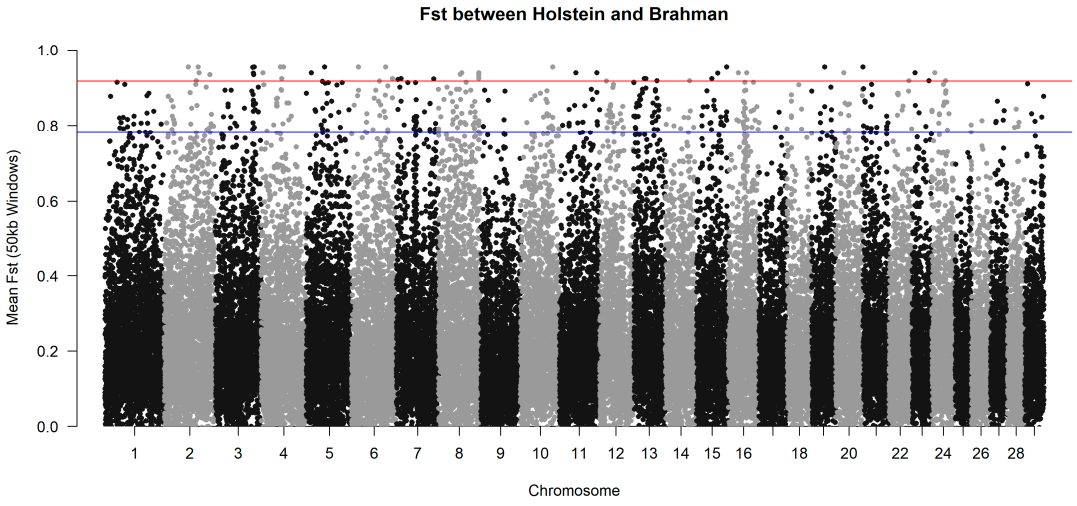

Figure 6. Manhattan plot depicting the genetic differentiation (FST) between the Holstein and Brahman cattle breeds.

Manhattan plot (Figure 7) showing the FST between Holstein and Jersey breeds. Despite belonging to *Bos taurus* breeds, the plot indicates moderate differentiation, with notable peaks on chromosomes 6, 12, and 17, which may be because of breed-specific traits, such as milk volume (Holstein) versus milk composition (Jersey).

 doi:10.20944/preprints202601.1602.v1

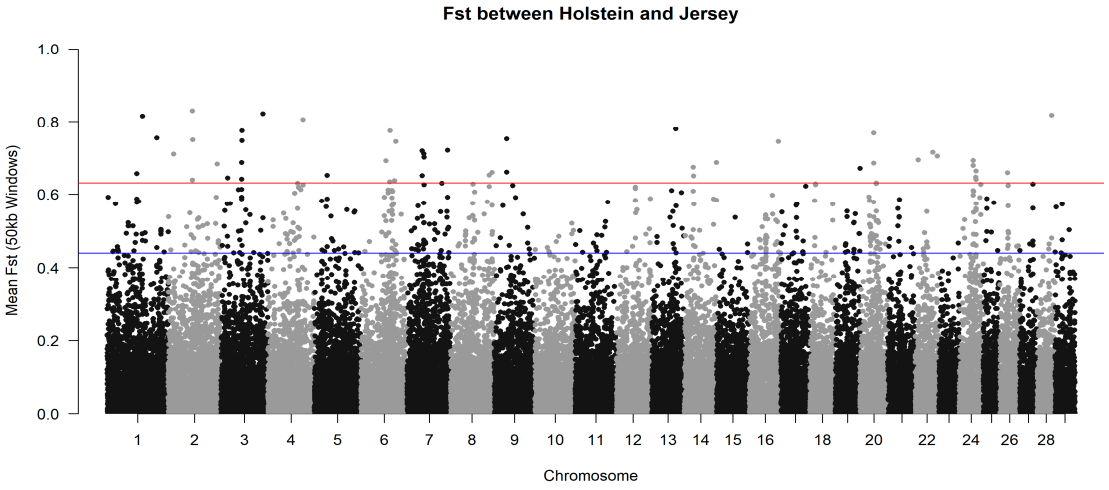

Figure 7. Manhattan plot depicting the genetic differentiation (FST) between the Holstein and Jersey cattle breeds.

Uniform high FST values across all autosomes can be seen in the Manhattan plot of Jersey and Brahman (Figure 8). With many exceeding the 99th and 99.9th percentile thresholds, it indicates a deep evolutionary split. High differentiation value reflects their distinct subspecies origins, different environmental pressures, and divergent selection for production traits.

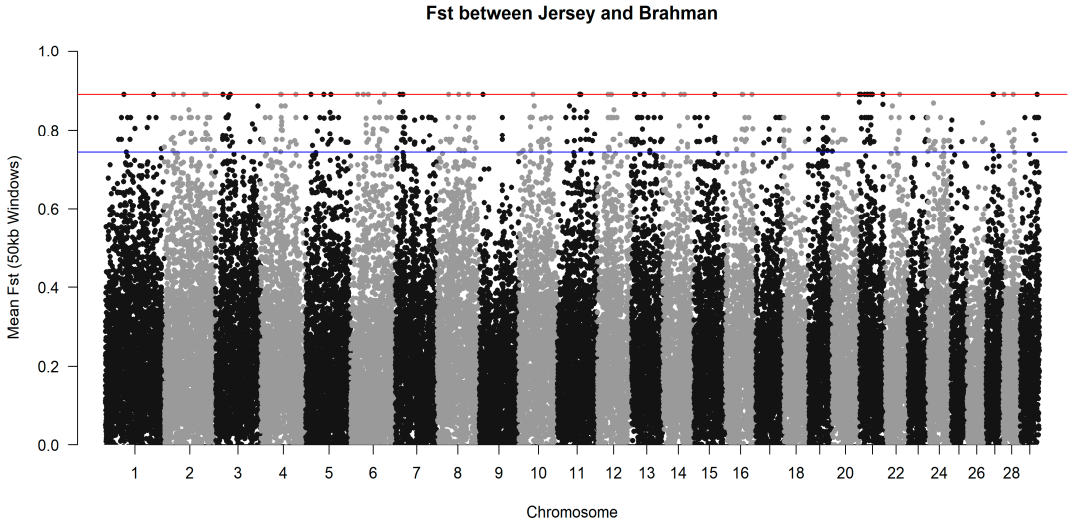

Figure 8. Manhattan plot showing the genetic differentiation (FST) between the Jersey and Brahman cattle breeds.

Manhattan plot (Figure 9) showing FST between Sahiwal and Holstein breeds. This FST distribution resembles closely with that of Cholistani-Holstein, with high genome-wide differentiation and peaks exceeding thresholds. This strong divergence reflects the contrast between a heat-adapted *Bos indicus* and a temperate *Bos taurus* breed.

 doi:10.20944/preprints202601.1602.v1

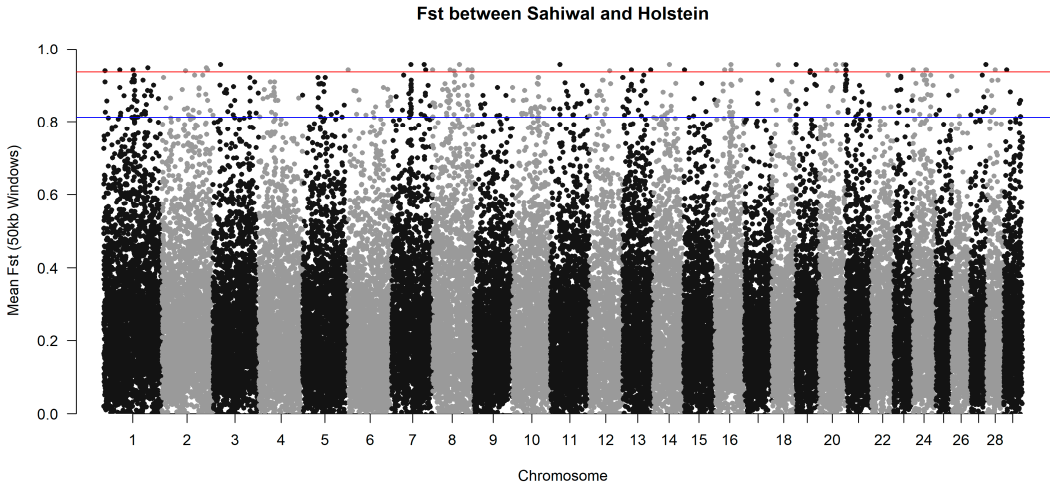

Figure 9. Manhattan plot illustrating the genetic differentiation (FST) between the Sahiwal and Holstein cattle breeds.

Manhattan plot (Figure 10) showing FST between Sahiwal and Brahman reveals moderate genome-wide FST values, with prominent peaks on chromosomes 1,13, and 22 exceeding the 99.9th percentile. Because of their origin from the *Bos indicus* background, differentiation is moderate.

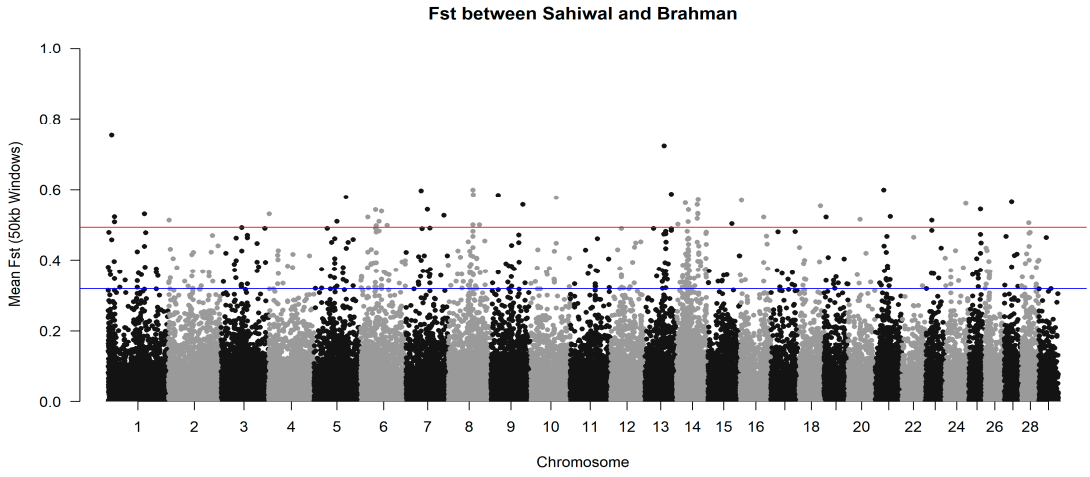

Figure 10. Manhattan plot illustrating the genetic differentiation (FST) between the Sahiwal and Brahman cattle breeds.

Manhattan Plot (Figure 11) between Sahiwal and Cholistani breeds revealed very low FST values, with only a few isolated peaks above the 99.9th percentile. The genetic relationship is consistent with their geographic proximity and shared evolutionary history.

 doi:10.20944/preprints202601.1602.v1

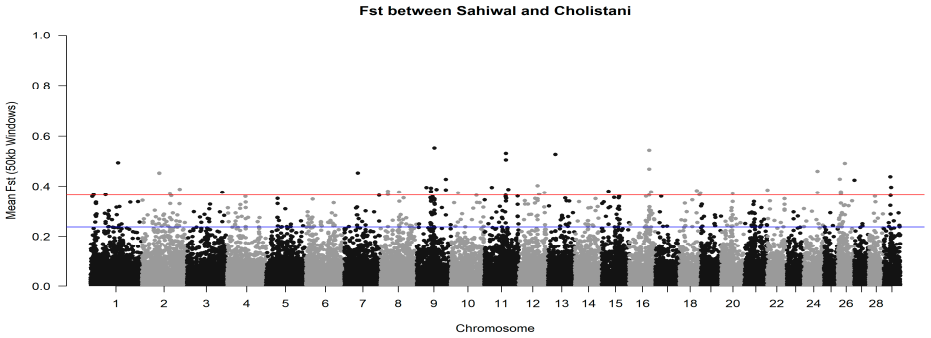

Figure 11. Manhattan plot depicting the genetic differentiation (FST) between the Sahiwal and Cholistani cattle breeds.

In (Figure 12), a Manhattan plot showing FST between Sahiwal and Jersey breeds. The plot shows high genome-wide FST values, as it was for Sahiwal and Holstein breeds. Because of deep evolutionary divergence between Bos indicus and Bos taurus, a significant proportion of genomic windows exceed both thresholds.

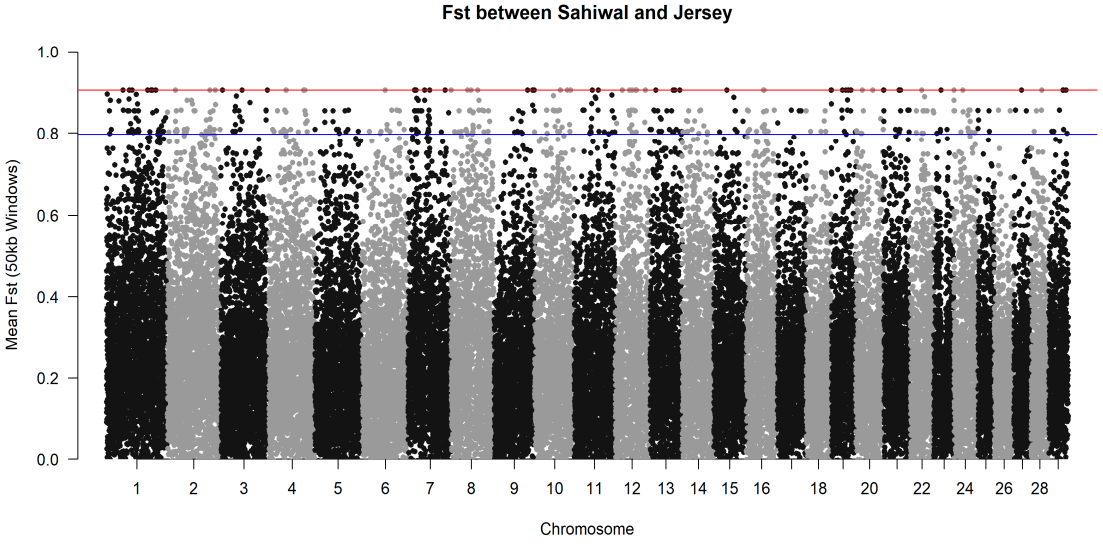

Figure 12. Manhattan plot showing the genetic differentiation (FST) between the Sahiwal and Jersey cattle breeds.

In this study, the SNP BovineHD2000011154 was identified as a highly polymorphic marker in the Sahiwal population (MAF = 0.3846) and was initially annotated as being located within the Prolactin Receptor (*PRLR*) gene on chromosome 20 (Table 1). The *PRLR* gene is a critical candidate for heat tolerance, famously housing the "Slick" mutation, making this finding highly relevant. To validate and further characterize the genomic context of this SNP, its position at **39,241,426 bp** was examined using the NCBI Genome Data Viewer against the latest bovine reference assembly (ARS-UCD1.3).

Table 1. Breed-specific polymorphic SNPs in PRLR and key heat stress–related genes in cattle.

Breed	Gene	Chr	SNP ID	Position (bp)	MAF	A1	A2
BRM	HSPB1	25	BovineHD2500011148	39,927,192	0.5	A	G
BRM	HSPB1	25	BovineHD2500011139	39,898,383	0.45	G	A
BRM	HSPB1	25	BovineHD2500011129	39,878,947	0.4	A	G

 doi:10.20944/preprints202601.1602.v1

BRM	HSF1	16	BovineHD1600013215	47,853,727	0.45	C	A
BRM	ATP2B1	3	BovineHD0300003806	11,513,093	0.45	A	G
BRM	PRLR	20	BovineHD2000011185	39,315,881	0.4	A	G
BRM	PRLR	20	BovineHD2000011129	39,192,015	0.35	G	A
BRM	PRLR	20	BovineHD2000011150	39,235,867	0.3	G	A
BRM	PRLR	20	BovineHD2000011154	39,241,426	0.3	C	A
BRM	FGF5	16	BovineHD1600002477	9,060,400	0.3	A	G
BRM	FGF5	16	BovineHD1600002485	9,102,200	0.25	A	G
BRM	HSPA1A	23	BovineHD2300007134	25,626,467	0.25	G	A
BRM	HSPA1A	23	BTA-118887-no-rs	25,642,674	0.1	A	G
CL	HSF1	16	BovineHD1600013215	47,853,727	0.5	C	A
CL	HSPB1	25	BovineHD2500011129	39,878,947	0.5	A	G
CL	HSPB1	25	BovineHD2500011139	39,898,383	0.38	G	A
CL	HSPB1	25	BovineHD2500011148	39,927,192	0.38	A	G
CL	ATP2B1	3	BovineHD0300003806	11,513,093	0.46	A	G
CL	PRLR	20	BovineHD2000011129	39,192,015	0.38	G	A
CL	PRLR	20	BovineHD2000011134	39,202,912	0.31	A	G
CL	PRLR	20	BovineHD2000011154	39,241,426	0.27	C	A
CL	PRLR	20	BovineHD2000011185	39,315,881	0.23	A	G
CL	FGF5	16	BovineHD1600002477	9,060,400	0.15	A	G
CL	FGF5	16	BovineHD1600002485	9,102,200	0.15	A	G
CL	HSPA1A	23	BTA-118887-no-rs	25,642,674	0.38	A	G
HOL	HSF1	16	BovineHD1600013215	47,853,727	0.44	C	A
HOL	ATP2B1	3	BovineHD0300003806	11,513,093	0.43	A	G
HOL	FGF5	16	BovineHD1600002485	9,102,200	0.42	A	G
HOL	FGF5	16	BovineHD1600002477	9,060,400	0.37	A	G
HOL	PRLR	20	BovineHD2000011185	39,315,881	0.4	A	G
HOL	PRLR	20	BovineHD2000011129	39,192,015	0.39	G	A
HOL	PRLR	20	BovineHD2000011154	39,241,426	0.38	C	A
HOL	PRLR	20	BovineHD2000011134	39,202,912	0.36	A	G
HOL	PRLR	20	BovineHD2000011150	39,235,867	0.35	G	A
HOL	HSPB1	25	BovineHD2500011139	39,898,383	0.41	G	A
HOL	HSPB1	25	BovineHD2500011129	39,878,947	0.29	A	G
HOL	HSPB1	25	BovineHD2500011148	39,927,192	0.17	A	G
HOL	HSPA1A	23	BovineHD2300007134	25,626,467	0.27	G	A
JER	FGF5	16	BovineHD1600002485	9,102,200	0.5	A	G
JER	ATP2B1	3	BovineHD0300003806	11,513,093	0.45	A	G
JER	PRLR	20	BovineHD2000011129	39,192,015	0.4	G	A
JER	PRLR	20	BovineHD2000011185	39,315,881	0.35	A	G
JER	HSF1	16	BovineHD1600013215	47,853,727	0.35	C	A
JER	HSPB1	25	BovineHD2500011129	39,878,947	0.35	A	G
JER	HSPB1	25	BovineHD2500011139	39,898,383	0.3	G	A
JER	HSPA1A	23	BovineHD2300007134	25,626,467	0.25	G	A
SW	ATP2B1	3	BovineHD0300003806	11,513,093	0.46	A	G
SW	HSF1	16	BovineHD1600013215	47,853,727	0.46	C	A
SW	HSPB1	25	BovineHD2500011139	39,898,383	0.42	G	A

 doi:10.20944/preprints202601.1602.v1

SW	HSPB1	25	BovineHD2500011148	39,927,192	0.38	A	G
SW	PRLR	20	BovineHD2000011129	39,192,015	0.38	G	A
SW	PRLR	20	BovineHD2000011154	39,241,426	0.38	C	A
SW	PRLR	20	BovineHD2000011185	39,315,881	0.31	A	G
SW	PRLR	20	BovineHD2000011134	39,202,912	0.23	A	G
SW	PRLR	20	BovineHD2000011150	39,235,867	0.15	G	A
SW	FGF5	16	BovineHD1600002477	9,060,400	0.35	A	G
SW	FGF5	16	BovineHD1600002485	9,102,200	0.31	A	G
SW	HSPA1A	23	BTA-118887-no-rs	25,642,674	0.23	A	G

The browser visualization confirmed that this position does indeed fall within a large intron of the *PRLR* gene. However, it also revealed a more precise annotation: the SNP is located within the gene body of a smaller, nested gene known as *TTC23L* (*Tetratricopeptide Repeat Domain 23 Like*). The location within a *PRLR* intron suggests a possible role in regulating *PRLR* expression or splicing; its concurrent location within *TTC23L* raises the possibility that it may also influence the function of this less-characterized gene. Therefore, BovineHD2000011154 represents a variant with potential dual-locus effects. This finding refines our understanding of the SNP's potential function. Given its high variability in Sahiwal cattle and its strategic position within the functionally significant *PRLR* locus, this SNP remains a high-priority candidate for future association studies aimed at dissecting the genetic architecture of thermotolerance and adaptation in *Bos indicus* cattle.

This particular finding, centered on the SNP BovineHD2000011129, tells a fascinating story about the genetic complexity of adaptation in Sahiwal cattle. When we first analyzed our data, this SNP was flagged as being located within the famous Prolactin Receptor (PRLR) gene on chromosome 20, a gene well-known for its connection to lactation and the "Slick" hair coat that helps cattle stay cool.

However, when we put this exact location, 39,192,015, under the microscope using the NCBI Genome Viewer (as shown in the image), we get a much more detailed and exciting picture. It turns out this SNP is in a classic "gene-within-a-gene" scenario. It resides within a large non-coding region (an intron) of the PRLR gene; this exact location is also part of a smaller, nested gene called *DNAJC21*.

This is where the story gets exciting. DNAJC21 is part of the Hsp40 family of proteins, which act as "co-chaperones." They are essential helpers for the major heat shock proteins (like Hsp70) that protect a cell's machinery from being damaged by heat stress.

So, what we have is a single genetic marker with a potential "two-for-one" deal in biological relevance: 1) It sits within the broader regulatory landscape of PRLR, a key gene for large-scale adaptation. 2) It is also located directly within DNAJC21, a gene specifically involved in the cellular stress response.

What makes this SNP a top-tier candidate for further study is its high variability in the Sahiwal population. With a Minor Allele Frequency (MAF) of 0.3846, both the 'G' and 'A' versions of this SNP are very common. This means the Sahiwal breed maintains a rich genetic toolkit at a genomic hotspot that could potentially influence two distinct but complementary pathways for dealing with heat. This high level of diversity at such a functionally important location makes BovineHD2000011129 an exceptionally strong candidate for future studies aiming to pinpoint the specific genetic variants that give Sahiwal cattle their renowned resilience.

Discussion

In this study, five prominent cattle breeds, Sahiwal, Cholistani, Holstein Friesian, Jersey, and Brahman, were investigated using high-density (HD) SNP genotyping to explore genetic structure and identify regions under selection. These breeds represent a blend of indigenous and exotic cattle populations, including both dairy and beef breeds, which are critical to livestock productivity in tropical and subtropical environments. The combination of Pairwise FST, Principal Component, and

 doi:10.20944/preprints202601.1602.v1

structure analysis highlights the strong genetic distinction between *Bos indicus* and *Bos taurus*, plus finer-scale differentiation among closely related populations.

The PCA results depicted a clear separation between the European and Asian breeds, with PC1 accounting for over one-third of the total variation. The variation reflects the deep evolutionary divergence between both breeds, which has been well documented in previous genomic studies of cattle diversity [18,19]. Whereas, PC2 shows 7% variation but within the *Bos indicus* group, which indicates breed-specific selection and adaptation patterns [20]. This pattern matches the finding of [21], who reported a clear difference between *Bos indicus* and *Bos taurus* groups. Further, PCA analysis comparing only Cholistani and Sahiwal breeds highlighted **moderate genetic differentiation**. This observation aligns with findings by [19], who confirmed these patterns with HD SNP data. These two breeds, although genetically related, have diverged slightly due to geographical separation and differing environmental pressures in their respective regions along the Indus River basin.

The structure analysis findings in this study also relate to past studies. At K=2, the major taurine-indicine division was evident, which aligns with the results of [22], who observed the same primary genetic division in Iraqi cattle. At higher K values, our analysis resolved each breed into distinct clusters, reflecting the genetic uniqueness maintained. Through selective breeding and limited recent admixture, comparable patterns were reported by [23]. He finds optimal clustering of separated indigenous breeds according to their adaptations as reported in [21]. Revealed the finer substructures corresponding to individual breeds.

To uncover the regions of the genome under positive selection, pairwise FST analysis was performed across all SNPs. The FST analysis provided clear insights into the genetic relationship and differentiation patterns among the five cattle breeds studied: Sahiwal, Cholistani, Brahman, Holstein, and Jersey. In this study, FST values were calculated by comparing heat-tolerant *indicine* breeds (excluding Sahiwal and Jersey in some analyses for clarity) against non-heat-tolerant breeds like Holstein Friesian and Jersey, following the groupings suggested by [19]. Consistent with expectations, the lowest genetic difference was between the *Bos indicus* group Cholistani and Sahiwal breeds because of their shared geographic origin, overlapping breeding histories, and similar adaptive pressures likely explain this close genetic affinity. A similar finding was reported with low differentiation among locally adapted breeds with shared origins in the Brazilian indicine cattle population, while the natural selection and historical gene flow have maintained high within-breed variability [24].

Moderate differentiation was found between indicine breeds from different origins, such as Sahiwal-Brahman or Cholistani-Brahman. Although these breeds belong to the indicine group, but still showed differentiations because of the adaptation to the different environments. The Manhattan plots for these breed comparisons revealed distinct genomic peaks, particularly on chromosomes 1,13, and 22, which may harbor genes related to thermotolerance, disease resistance, or reproductive performance. This study is in line with the findings from Mexican Coreno creole cattle, where the genetic difference was not that high, but certain loci linked to candidate genes such as *GOT1* and *NCAD* were implicated in heat stress response and homeostatic processes [25].

In our study, the highest values were obtained when the comparison was done between *Bos indicus* (Cholistani, Sahiwal, and Brahman) and the *Bos taurus* group (Holstein Friesian and Jersey). The widespread and genomic differentiations with many windows exceeding the 99.9th percentile threshold reflect their deep evolutionary split and strong artificial selection for different production systems for high yield, like in taurine or the heat-adapted traits in indicines. This finding aligns with the findings of [24], who performed FST on Brazilian zebu and taurine cattle using SNP markers, and the FST values between two cattle groups were substantial and the values were high, as in this study.

The FST results showed a clear pattern for genetic differentiation among breeds. Those who live in the same local environment show low FST value, same breed but different environment showed moderate values, whereas when it comes to the different breeds, peak values cross the percentile threshold in excess. These findings underscore the importance of conserving genetic diversity within indigenous breeds, which harbor valuable adaptive traits. Strategies in developing crossbreds that

 doi:10.20944/preprints202601.1602.v1

maintain productivity under projected climate conditions, this genetic differentiation will be extremely important.

Conclusion

This study provides a genome-wide evaluation of genetic differentiation and selection signatures underlying heat stress adaptation in five cattle breeds representing contrasting evolutionary origins and climatic adaptation. Population structure analyses confirmed a clear separation between Bos indicus and Bos taurus breeds, with low differentiation between the indigenous Sahiwal and Cholistani cattle and moderate divergence of Brahman reflecting shared indicine ancestry with distinct breeding histories. Genome-wide F_{ST} analyses revealed extensive differentiation between indicine and taurine breeds and identified multiple genomic regions exhibiting strong signals of selection. Notably, several high-confidence selection signatures localized to the PRLR locus and key heat shock protein–related genes, including HSPB1, HSPA1A, HSF1, ATP2B1, and FGF5, highlighting their central role in endocrine regulation, cellular stress response, and thermotolerance. The presence of highly polymorphic variants within intronic and nested gene regions of PRLR suggests potential regulatory mechanisms contributing to adaptive resilience under heat stress. Collectively, these findings strengthen the evidence that heat tolerance in cattle is shaped by coordinated selection on endocrine and cellular stress pathways. The genomic regions and candidate loci identified in this study provide a valuable foundation for future functional validation and for the development of marker-assisted and genomic selection strategies aimed at improving thermal resilience while maintaining productivity. Preservation and strategic utilization of adaptive genetic variation within indigenous cattle populations will be critical for sustaining dairy production in hot environments under ongoing climate change.

Author Contributions: H.M. conceived and designed the study, supervised the analyses, and drafted the manuscript. M.A.-A. contributed to study design, interpretation of results, and critical revision of the manuscript. W.A. assisted with data acquisition and coordination with industry partners. H.K. performed data curation, statistical analyses, and figure preparation. W.A.K. contributed to genomic interpretation and manuscript editing. J.-J.K. provided expertise in population genomics and selection signature analysis and contributed to interpretation of the results. M.B.B.M. assisted with data analysis, validation of results, and manuscript drafting. All authors read and approved the final manuscript.

Funding Information: This study utilized genotype data generated under a research project funded by the Higher Education Commission (HEC) of Pakistan (NRPU No. 16844). No additional external funding was received for data analysis or manuscript preparation.

Data Availability: The genotype data analyzed in this study are not publicly available due to institutional and ethical restrictions. Data may be made available from the corresponding author upon reasonable request and subject to approval by the data-owning institution.

Acknowledgments: The authors acknowledge the Higher Education Commission (HEC) of Pakistan for facilitating access to genotype data. We thank the Livestock and Dairy Development Department (L&DD), Punjab, and other collaborating institutions for their support in sample collection and data generation. The authors also appreciate the technical support provided by colleagues at the Department of Animal Breeding and Genetics, University of Veterinary and Animal Sciences (UVAS), Lahore. Maxim Agri (Pvt.) Ltd. is gratefully acknowledged for providing samples and logistical support.

Conflicts of Interest Statement: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Verma, K. K., Song, X.-P., Kumari, A., Jagadesh, M., Singh, S. K., Bhatt, R., Singh, M., Seth, C. S., & Li, Y.-R. (2025). *Climate change adaptation: Challenges for agricultural sustainability*. **Plant, Cell & Environment, 48**, 2522–2533. https://doi.org/10.1111/pce.15078.
2. Godde, C. M., Mason-D'Croz, D., Mayberry, D. E., Thornton, P. K., & Herrero, M. (2021). *Impacts of climate change on the livestock food supply chain: A review of the evidence*. **Global Food Security, 28**, 100488. https://doi.org/10.1016/j.gfs.2020.100488.
3. Habeeb, A. A. M. (2020). *Deterioration effects of heat stress on farm animals performance in tropical and subtropical regions*. **World Journal of Biology Pharmacy and Health Sciences, 4**(2), 007–025. https://doi.org/10.30574/wjbphs.2020.4.2.0088.
4. Usman, M., Ali, A., Rosak-Szyrocka, J., Pilař, L., Baig, S. A., Akram, R., & Wudil, A. H. (2023). Climate change and livestock herders' wellbeing in Pakistan: Does the nexus of risk perception, adaptation and their drivers matter? Heliyon, 9(6), e16983. https://doi.org/10.1016/j.heliyon.2023.e16983.
5. Hansen, P. J. (2004). *Physiological and cellular adaptations of zebu cattle to thermal stress*. **Animal Reproduction Science, 82–83**, 349–360. https://doi.org/10.1016/j.anireprosci.2004.04.011.
6. Beatty D.T., Barnes A., Taylor E., Pethick D., McCarthy M., Maloney S.K. Physiological responses of Bos taurus and Bos indicus cattle to prolonged, continuous heat and humidity. J. Anim. Sci. 2006;84:972–985. doi: 10.2527/2006.844972x.
7. Davila K.M.S., Hamblen H., Hansen P.J., Dikmen S., Oltenacu P.A., Mateescu R.G. Genetic parameters for hair characteristics and core body temperature in a multibreed Brahman—Angus herd1. J. Anim. Sci. 2019;97:3246–3252. doi: 10.1093/jas/skz188.
8. Dikmen S., Mateescu R.G., Elzo M.A., Hansen P.J. Determination of the optimum contribution of Brahman genetics in an Angus-Brahman multibreed herd for regulation of body temperature during hot weather. J. Anim. Sci. 2018;96:2175–2183. doi: 10.1093/jas/sky133.
9. Worku D, Hussen J, De Matteis G, Schusser B and Alhussien MN (2023) Candidate genes associated with heat stress and breeding strategies to relieve its effects in dairy cattle: a deeper insight into the genetic architecture and immune response to heat stress. Front. Vet. Sci. 10:1151241. doi: 10.3389/fvets.2023.1151241.
10. Gomez-Pastor, R., Burchfiel, E. T., & Thiele, D. J. (2018). *Regulation of heat shock transcription factors and their roles in physiology and disease*. **Nature Reviews Molecular Cell Biology, 19**(1), 4–19. https://doi.org/10.1038/nrm.2017.73.
11. Fariha, Shabbir, M. S., Anwar, S., Azmat, H., Bin Majeed, M. B., & Kaul, H. (2025). *Comparative analysis of heat stress response in Holstein Friesian and Sahiwal cattle through HSP70 gene expression and promoter DNA methylation*. **Animal Genetics**. https://doi.org/10.1111/age.70048.
12. Tiwari, M., Gujar, G., Shashank, C. G., & Ponsuksili, S. (2024). *Selection signatures for high altitude adaptation in livestock: A review*. **Gene, 927**, 148757. https://doi.org/10.1016/j.gene.2024.148757.
13. Niu P, Li X, Wang X, Qu H, Chen H, Huang F, Hu K, Fang D, Gao Q. Population Genetic Structure, Historical Effective Population Size, and Dairy Trait Selection Signatures in Chinese Red Steppe and Holstein Cattle. Animals (Basel). 2025 Aug 27;15(17):2516. doi: 10.3390/ani15172516. PMID: 40941311; PMCID: PMC12427281.
14. Cheruiyot EK, Haile-Mariam M, Cocks BG and Pryce JE (2022) Improving Genomic Selection for Heat Tolerance in Dairy Cattle: Current Opportunities and Future Directions. Front. Genet. 13:894067. doi: 10.3389/fgene.2022.894067.
15. Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). *Second-generation PLINK: Rising to the challenge of larger and richer datasets*. **GigaScience, 4**(1), Article 7. https://doi.org/10.1186/s13742-015-0047-8.
16. Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evol. 38(6): 1358–70.
17. Zhu Z, Wang Y, Zhou X, Yang L, Meng G, Zhang Z. 2020. SWAV: a web-based visualization browser for sliding window analysis. Sci Rep. 10(1): 149. https://doi.org/10.1038/s41598-019-57038-x.
18. Lewis J, Abas Z, Dadousis C, Lykidis D, Paschou P, Drineas P. 2011. Tracing cattle breeds with principal components analysis ancestry informative SNPs. PloSone. 6(4): e18007.

 doi:10.20944/preprints202601.1602.v1

19. Mustafa H, Khan WA, Sonstegard T, Li Y, Ain NU, Ajmal A. 2018. Genome-Wide Identification of Natural Selection Footprints in *Bos Indicus* Using Principal Component Analysis. Adv Life Sci. 5(2): 67-72.
20. Bahbahani H, Clifford H, Wragg D, Mbole-Kariuki MN, Van Tassell C, Sonstegard, T Hanotte O. 2015. Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. Sci rep. 5(1): 11729. http://doi.org/10.1038/srep11729.
21. Edea Z, Bhuiyan MSA, Dessie T, Rothschild MF, Dadi H, Kim KS. 2015. Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. Anim. 9(2): 218-226.
22. Alshawi A, Essa A, Al-Bayatti S, Hanotte O. 2019. Genome analysis reveals genetic admixture and signature of selection for productivity and environmental traits in Iraqi cattle. Front genet.10: 609. http://doi.org/10.3389/fgene.2019.00609.
23. Goitom S, Gicheha M G, Njonge F K, Kiplangat N. 2019. Genome-wide genetic diversity, population structure and admixture analysis in Eritrean Indigenous Cattle. S Afr J Anim Sci. 49(6): 1083-1092. https://doi.org/10.1093/nar/17.20.8390.
24. Campos BM, do Carmo AS, do Egito AA, da Mariante AS, do Albuquerque M SM, de Gouveia JJS, Carneiro PLS. 2017. Genetic diversity, population structure, and correlations between locally adapted zebu and taurine breeds in Brazil using SNP markers. Trop Anim Health Prod. 49(8): 1677-1684.
25. Luna-Azuara C G, Montano-Bermudez, M., Calderon-Chagoya R, Rios-Utrera A, Martinez-Velazquez G, Vega-Murillo V E. 2024. Genetic diversity of SNPs associated with candidate genes for heat stress in Coreño Creole cattle in Mexico. Trop Anim Health Prod. 56(2): 71. http://doi.org/10.1007/s11250-024-03917-z.