

RESEARCH ARTICLE

Genome-wide identification of candidate breed-specific markers and selection signatures in Kalmyk cattle



Zanda Bochkaeva¹ , Aleksandra Abdelmanova² , Tatiana Deniskova² , Victoria Ubushieva¹ , Altana Ubushieva¹ , Nadezhda Chimidova¹ , Arslang Khakhlinov¹ , Baatr Bolaev¹ , Badma Salaev¹ , and Natalia Zinovieva² 

1. Regional Center of Science and Production, Kalmyk State University named after B.B. Gorodovikov, 11 Pushkin Street, Elista 358000, Republic of Kalmykia, Russian Federation.
2. L.K. Ernst Federal Research Center for Animal Husbandry, Dubrovitsy, Podolsk, Moscow Oblast 142132, Russia.

ABSTRACT

Background and Aim: Kalmyk cattle (KALMYK) represent an indigenous Russian beef breed with valuable adaptive and productive characteristics under harsh continental steppe conditions. However, previous genomic studies have involved relatively small populations, and candidate markers that distinguish KALMYK from both specialized beef and steppe cattle remain insufficiently characterized. This study aimed to characterize the population structure and genetic diversity of KALMYK and identify candidate breed-specific single-nucleotide polymorphisms (SNPs), genomic regions under putative selection pressure, and associated candidate genes using genome-wide genotyping.

Materials and Methods: Genome-wide genotypes were obtained from 100 purebred KALMYK using the Illumina BovineSNP50 v3 BeadChip and compared with Hereford ($n = 27$), Kazakh White-Headed ($n = 93$), Red Steppe ($n = 45$), and Mongolian ($n = 47$) cattle. Across platforms, 44,774 common SNPs were retained before quality control. The final dataset comprised 312 animals and 30,839 SNPs, with 27,932 linkage disequilibrium-pruned SNPs used for population structure analyses. Genetic diversity, population structure, differentiating SNPs, runs of homozygosity, F_{ST} , and cross-population extended haplotype homozygosity were evaluated.

Results: KALMYK showed the closest genetic relationship with Mongolian cattle ($F_{ST} = 0.019$). Two comparisons, Kalmyk versus specialized beef breeds and Kalmyk versus steppe cattle, each identified 31 highly differentiated SNPs. After accounting for one shared locus, 61 candidate SNPs were retained. Fifty-one genomic regions across 19 chromosomes contained 202 genes, including a shared region on *Bos taurus* autosome (BTA)12 harboring *RNASEH2B*. Other prioritized genes included *USP18*, *C1QTNF6*, *SSTR3*, *RAC2*, and *SYT10*. Three islands of homozygosity were detected on BTA5, BTA7, and BTA13. The mean genomic inbreeding coefficient was 0.05 in KALMYK compared with 0.133 in specialized beef breeds.

Conclusion: This study provides a genome-wide framework for characterizing the KALMYK gene pool and proposes 61 candidate SNPs for future breed authentication and genetic monitoring. Because we did not perform phenotypic association or independent validation, these markers remain candidates that require validation in independent purebred and crossbred populations before practical implementation.

Keywords: breed authentication, cattle genomics, genetic diversity, genome-wide genotyping, Kalmyk cattle, selection signatures, single-nucleotide polymorphisms, SNP markers.

INTRODUCTION

Preserving the genetic diversity of national farm animal resources is essential for ensuring food security and the sustainable development of livestock production systems [1]. This issue has become increasingly important under changing climatic conditions, which have intensified interest in indigenous breeds that can maintain productivity in harsh environments with relatively low external inputs. Kalmyk cattle (KALMYK) are among the oldest indigenous Russian cattle breeds and represent an important beef genetic resource. The breed is believed to have developed from local cattle maintained by nomadic populations under the harsh climatic conditions of China and Central Asia and introduced from Western Mongolia in the 17th century [2]. Prolonged exposure to these environments and strong natural selection are considered to have contributed to the development of its

Corresponding Author: Zanda Bochkaeva

E-mail: zbochkaeva.npc@mail.ru

Received: 10-06-2026, **Accepted:** 16-09-2026, **Published online:** 09-10-2026

Co-authors: AA: preevetic@mail.ru; TD: horarka@yandex.ru; VU: vicki_93g@mail.ru; AU: ameli-altanas@mail.ru; NC: nadezhdashimidova@yandex.ru; AK: arspeople@mail.ru; BB: kanur64@mail.ru; BS: uni@kalmsu.ru; NZ: n_zinovieva@mail.ru

How to cite: Bochkaeva Z, Abdelmanova A, Deniskova T, Ubushieva V, Ubushieva A, Chimidova N, et al. Genome-wide identification of candidate breed-specific markers and selection signatures in Kalmyk cattle. Vet World, 2026, 19(10): 4498–4528.

Copyright: Bochkaeva et al. This article is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).



adaptive characteristics.

KALMYK are adapted to an extreme continental climate, with temperatures ranging from approximately -40°C to $+50^{\circ}\text{C}$. Previous genome-wide research has provided initial insights into the genetic basis underlying adaptation of this breed to the environmental conditions of its historical habitat [3]. In addition to adaptation, KALMYK possess economically important productive and reproductive characteristics, including favorable beef production and meat quality traits [4]. Mature cows attain live weights of approximately 450–480 kg and bulls 800–950 kg, while an average calving rate of approximately 90% has been reported [5]. These productive and reproductive characteristics, along with the breed's ability to perform under extensive production conditions, contribute to its importance as an indigenous beef genetic resource.

Phylogenetic investigations indicate a Turano-Mongolian origin of KALMYK. Their genetic composition is hypothesized to include a major ancestral component related to Mongolian cattle (MONG), along with a smaller contribution from other Eurasian taurine populations [2, 6]. This distinctive genetic background may contribute to the breed's adaptation to pronounced seasonal climatic variation, although the genomic architecture underlying many of these characteristics remains incompletely understood. Production studies have also demonstrated the economic potential of KALMYK. From 6 to 12 months of age, purebred Kalmyk bulls achieved an average daily gain of 750 g compared with 670 g in Angus $\times$ Kalmyk crossbred bulls while requiring substantially less feed per kilogram of gain [7]. Under intensive rearing from 8 to 12 months of age, Kalmyk bulls achieved an average daily gain of 1540 g, comparable to Russian Komola and Kazakh White-Headed cattle (KZWH), and showed favorable production and economic characteristics [8]. Although Aberdeen Angus bulls have demonstrated higher pre-slaughter live weight and slaughter yield in comparative studies, Kalmyk beef has shown favorable characteristics, including higher protein content, pH, and water-binding capacity and lower cooking losses [9]. Collectively, these adaptive, productive, and reproductive attributes emphasize the importance of conserving the authentic genetic diversity of KALMYK.

The development of high-throughput genotyping platforms that can simultaneously analyze tens of thousands of single-nucleotide polymorphisms (SNPs) has substantially expanded opportunities to characterize livestock genomic diversity. Genome-wide SNP data can be used to investigate population structure, detect genomic regions potentially affected by selection, and identify candidate genes associated with economically important and adaptive characteristics. For example, complementary analyses based on F_{ST} , runs of homozygosity (ROH) islands, and hapFLK have identified genomic regions under putative selection pressure and candidate genes in Kholmogory and Yaroslavl cattle [10]. Analysis using the Bovine GGP HD BeadChip demonstrated the preservation of substantial authentic genomic components in five Russian dairy cattle breeds: Russian Black-Pied, Istoben, Kholmogory, Tagil, and Yaroslavl [11]. Genome-wide investigations have also identified genes associated with feed efficiency, growth, and carcass characteristics in Red Angus cattle [12]. Candidate genes associated with metal ion binding, muscle growth and differentiation, and immunity, including *KCNK1*, *PDLIM5*, *CPXM2*, *CAPN14*, *MIR2285D*, *MYOF*, *PKDCC*, *FOXN3*, and *EHD3*, have been reported in Huaxi cattle [13]. Similarly, *UQCR11*, *DNAJC18*, *EGR1*, and *STING1* have been identified as candidate functional genes associated with thermogenesis and energy metabolism in indigenous Chinese cattle maintained under different climatic conditions [14]. Genomic regions and quantitative trait loci associated with resistance to blood-feeding insect bites have also been identified in West African cattle [15].

The selection of informative SNPs for breed-specific marker panels represents another important application of genome-wide genotyping, particularly for genetic certification, breed authentication, and conservation of valuable authentic genomic components. Using the Illumina BovineSNP50 BeadChip, breed-specific panels containing highly informative SNP markers have been developed for 10 local European cattle breeds [16]. Panels comprising approximately 5000 SNPs have been proposed for genetic certification of local cattle from Hainan Island [17], whereas approximately 200,000 SNPs have been used to characterize the genetic structure of endangered German Black-Pied cattle [18]. Furthermore, 5K and 10K SNP panels have been successfully validated for differentiation among Argentine cattle populations [19]. These studies demonstrate the potential of reduced, informative marker sets for population discrimination and genetic resource management.

Previous genome-wide investigations of KALMYK have provided valuable information on their population structure, genetic relationships, and loci associated with economically important characteristics. Kalmyk and Yakut cattle have been shown to differentiate from several other cattle populations, consistent with their Turano-Mongolian ancestry [6]. Three SNP markers, Hapmap46744BTA03982, ARSBFGLNGS12917, and LOC787924, have been associated with live weight in Kalmyk bulls [20]. The genomic contribution of KALMYK to the KZWH breed

has also been investigated [21], including evidence that a genomic region containing *SRD5A2* was inherited by KZWH cattle from Kalmyk ancestors [22]. KALMYK are particularly valuable because of their adaptation to harsh continental environments characterized by long, cold winters and hot, dry summers, which facilitates extensive pasture-based production [23]. Genetic heterogeneity has also been reported for markers in genes related to meat productivity, including *GH*, *LEP*, *TG*, and *CAPN1*, indicating opportunities for genetic improvement while conserving the breed's adaptive potential [24].

Despite these advances, the genomic characterization of KALMYK remains incomplete. Previous genome-wide studies included relatively small Kalmyk populations, generally comprising 12–24 animals [6, 20], which may limit representation of within-breed genomic diversity. Moreover, systematic identification of candidate breed-differentiating SNPs using a sufficiently large purebred Kalmyk population and simultaneous comparisons with both specialized beef and steppe cattle has not been adequately addressed. Previous investigations also have not combined breed differentiation with multiple complementary approaches to detect genomic regions under putative selection pressure within a unified analytical framework. Consequently, candidate SNPs suitable for subsequent development and validation of a Kalmyk breed authentication panel remain insufficiently characterized. A larger genome-wide analysis integrating population structure, genetic diversity, candidate differentiating SNPs, and complementary selection-signature approaches is therefore required to characterize the distinctive genomic features of this indigenous breed more comprehensively.

Therefore, this study aimed to comprehensively characterize the genomic architecture of KALMYK using genome-wide SNP genotyping of a comparatively large purebred population and a dual comparative framework involving specialized beef and steppe cattle. Specifically, the objectives were to (1) characterize the population structure and genetic diversity of KALMYK relative to the reference populations; (2) identify candidate SNP markers differentiating KALMYK from specialized beef breeds and steppe cattle; (3) identify genomic regions under putative selection pressure using multiple complementary genome-wide approaches; and (4) annotate and characterize candidate genes located within the identified genomic regions. By comparing KALMYK separately with specialized beef breeds and steppe cattle, the study was designed to identify complementary patterns of genomic differentiation associated with these contrasting reference groups. The resulting candidate markers and genomic regions are intended to provide a foundation for further independent validation and the potential future development of a marker panel for genetic characterization and authentication of KALMYK.

MATERIALS AND METHODS

Ethical approval

This study was conducted in accordance with applicable ethical and animal-welfare standards. All procedures complied with Directive 2010/63/EU of the European Parliament and of the Council on the protection of animals used for scientific purposes and with the Russian Federation national standards GOST 33044-2014 and GOST 34088-2017. The study procedures were conducted in accordance with the principles of the 3Rs (Replacement, Reduction, and Refinement).

The animal study protocol was reviewed and approved by the Commission on Ethics of Animal Experiments of the L.K. Ernst Federal Research Center for Animal Husbandry (Protocol No. 2, dated February 2, 2025). Blood collection was performed by professional veterinarians concurrently with routine veterinary procedures, after which all animals remained on their respective breeding farms.

Study period and location

The study was conducted from May to November 2025. The study involved purebred KALMYK maintained at breeding farms located in the Republic of Kalmykia, Russian Federation. Bulls were obtained from four breeding farms: Pik-Plus in Priyutnensky District, Sarpa in Ketchenerovsky District, Tsagan-Usn in Yashkulsky District, and the farm named after A. Chapchaev in Ketchenerovsky District. Heifers originated from Uralan in Priyutnensky District, Kirovsky in Yashkulsky District, Pervomaysky in Chernozemelsky District, and the Regional Center for Research and Production of Kalmyk State University named after B.B. Gorodovikov.

Study design

This genome-wide comparative study aimed to characterize the genetic diversity and population structure of KALMYK and identify candidate genomic markers and regions that differentiate the breed from specialized beef and steppe cattle populations. A total of 100 clinically healthy purebred KALMYK were included and compared with reference genotypes from Hereford, KZWH, Red Steppe, and MONG. The analytical workflow incorporated

genome-wide genotyping, quality control, population structure analysis, estimation of genetic differentiation, identification of differentiating SNPs, ROH analysis, XP-EHH analysis, and functional annotation of candidate genomic regions (Figure 1).

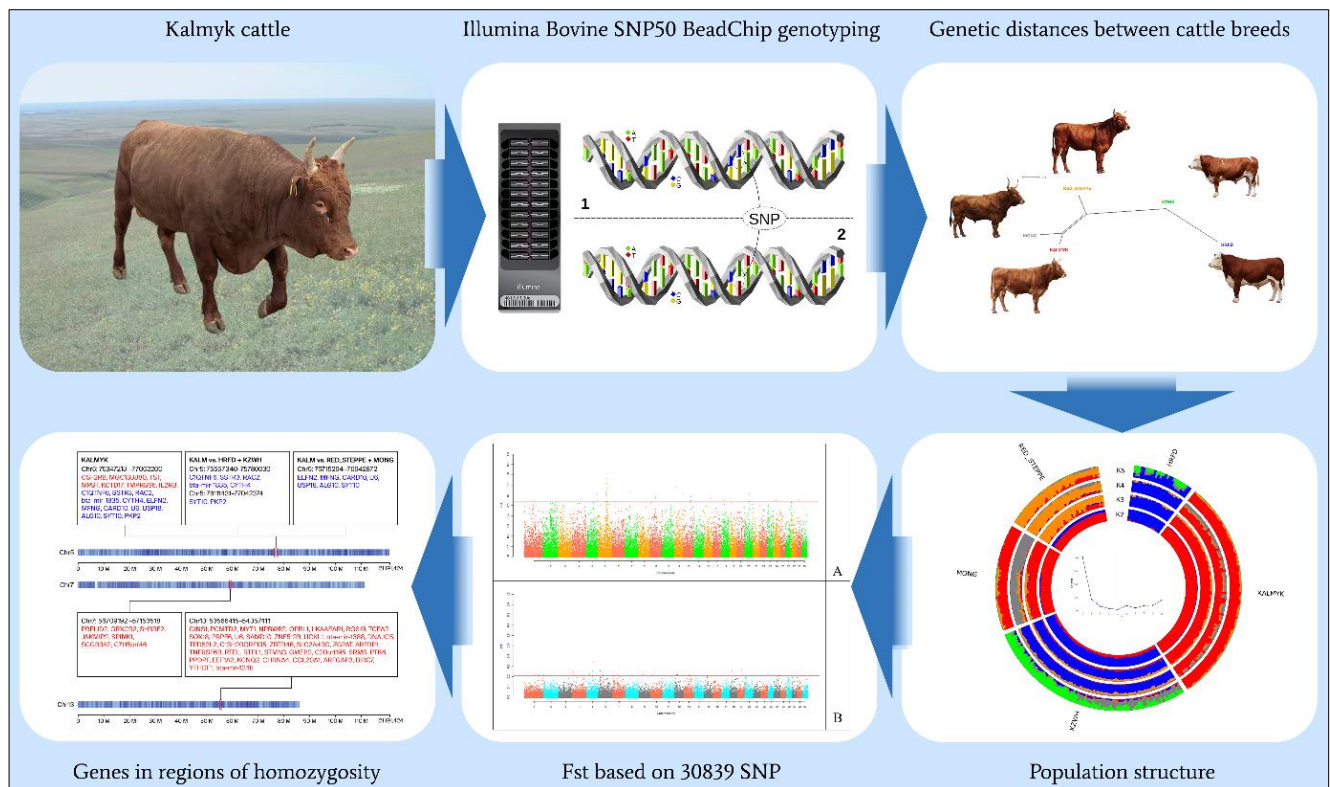


Figure 1: Graphical representation of the study design [Source: The figure was generated using AI Jay Flow].

Animals and sampling

Blood samples were collected from 100 clinically healthy purebred KALMYK, comprising 50 bulls aged 8 months and 50 heifers aged 2–3 years. The bulls originated from Pik-Plus, Sarpa, Tsagan-Usn, and the farm named after A. Chapchaev, whereas the heifers originated from Uralan, Kirovsky, Pervomaysky, and the Regional Center for Research and Production of Kalmyk State University named after B.B. Gorodovikov.

Farms and animals were selected using a stratified random-sampling approach to represent the principal geographical subpopulations of KALMYK within the Republic of Kalmykia. To reduce close relatedness, animals with known common parents within two generations were excluded before sampling. Purebred status was verified using breeding certificates and records in the State Breeding Register. The final sample comprised equal numbers of bulls and heifers. Because downstream genomic analyses were based on autosomal SNPs after exclusion of sex-linked markers, sex was not included as a covariate.

The animals were maintained on breeding farms under predominantly free-grazing steppe conditions. Blood collection was performed by professional veterinarians during routine veterinary procedures. Following sampling, all animals remained on their farms as members of the Kalmyk breeding population.

Approximately 4 mL of blood was collected from the jugular vein of each animal into sterile Ethylenediamine-tetraacetic acid vacuum tubes (Improvacuter; Guangzhou Improve Medical Instruments Co., Ltd., Guangzhou, China). Blood samples were stored at -20°C until DNA extraction.

DNA extraction and genome-wide genotyping

Genomic DNA was extracted from 200 μL of whole blood using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA concentration was determined using a Qubit 4 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), whereas purity was assessed using the A260/A280 ratio with a NanoDrop 2000 spectrophotometer. Samples showed A260/A280 ratios of 1.80–1.90 and DNA concentrations ≥ 50 ng/ μL .

Genome-wide genotyping of KALMYK was performed at the genetics laboratory of LLC Miratorg-Genetics using the Illumina BovineSNP50 v3 BeadChip (Illumina Inc., San Diego, USA). The resulting Kalmyk dataset comprised 100 animals (KALMYK, $n = 100$). Sample call rates ranged from 0.961 to 0.991, with most samples

exceeding 0.980; therefore, no Kalmyk samples were excluded using the predefined sample call rate threshold of <0.75.

Reference SNP profiles were obtained from the database of the L.K. Ernst Federal Research Center for Animal Husbandry for KZWH (n = 93), Red Steppe (RED_STEPPE, n = 45), and Mongolian (MONG, n = 16) cattle. Publicly available genotype datasets were additionally obtained for Hereford cattle (HRFD, n = 27) [25] and XilinGol MONG from present-day Inner Mongolia (MGX, n = 31) [26].

Preliminary analysis showed genetic similarity between MONG represented in the L.K. Ernst Federal Research Center DNA bank and the publicly available XilinGol MONG. These animals were therefore combined into a single Mongolian group (MONG, n = 47).

The combined dataset contained genotypes generated using three Illumina platforms: BovineSNP50 v1 (53,599 SNPs), BovineSNP50 v3 (53,218 SNPs), and BovineHD BeadChip (777,962 SNPs). Only the 44,774 SNPs shared among all three platforms were retained. Genotype imputation was not performed; all downstream analyses were based exclusively on the intersecting SNP set.

Processing of genotyping data and population structure analysis

Genotyping quality control included GenCall (GC) and GenTrain (GT) thresholds of 0.5 [27]. PLINK version 1.9 (<https://www.cog-genomics.org/plink/>) [28] was used for subsequent filtering. Samples with <75% successfully genotyped loci were excluded using `--mind 0.25`. SNPs with call rates <90% were removed using `--geno 0.1`, and SNPs with minor allele frequency <5% were removed using `--maf 0.05`. No duplicate samples were included. Mendelian-error assessment was not applicable because family trios were unavailable.

Cryptic relatedness was assessed using PLINK's `--genome` procedure. One individual from each pair with `PI_HAT > 0.2` was planned for exclusion; however, no pairs exceeded this threshold. After quality control, the final dataset comprised 312 animals and 30,839 SNPs: KALMYK (n = 100), HRFD (n = 27), KZWH (n = 93), RED_STEPPE (n = 45), and MONG (n = 47).

For analyses requiring reduced linkage disequilibrium (LD), additional pruning was performed using a sliding window of 50 SNPs, a step size of five SNPs, and an r^2 threshold of 0.5 (`--indep-pairwise 50 5 0.5`). A total of 2907 SNPs were removed. SNPs located on sex chromosomes and markers with unknown genomic positions were also excluded. SNP positions were originally based on the *Bos taurus* UMD3.1.1 assembly (GCF_000003055.6), whereas all genomic coordinates reported in this manuscript were converted to the *Bos taurus* ARS-UCD1.2 assembly (GCF_002263795.3). The resulting LD-pruned dataset contained 27,932 SNPs.

Genetic diversity was assessed separately within each breed using the R package *diveRsity* version 1.9.90 (<https://cran.r-project.org/web/packages/diveRsity/index.html>) [29]. Observed heterozygosity (H_O), unbiased expected heterozygosity (uH_E) [30], unbiased inbreeding coefficient (uF_{IS}) [31], and allelic richness (A_R), corrected for sample size [32], were calculated.

Genetic differentiation among breeds was assessed using principal component analysis (PCA), with visualization performed using the R package *ggplot2* version 3.5.0 [33]. Pairwise identical-by-state (IBS) genetic distances were used to construct Neighbor-Net networks using *SplitsTree* version 4.14.5 (<https://uni-tuebingen.de/en/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/algorithms-in-bioinformatics/software/splitstree/>) [34].

Pairwise F_{ST} values [31] were calculated using the R package *diveRsity* [29] and visualized using Neighbor-Net networks in *SplitsTree* version 4.14.5.

Population structure was additionally assessed using ADMIXTURE version 1.23 [35]. The optimal number of ancestral clusters (K) was determined by identifying the lowest cross-validation error across K = 1–10. LD pruning for ADMIXTURE was performed using the PLINK `--indep-pairwise 50 5 0.5` procedure. Visualization was performed using the R package *BITE* version 2.2.1 (<https://github.com/marcomilanesi/BITE>) [36].

Identification of breed-specific SNPs

Candidate differentiating SNPs distinguishing KALMYK from the other studied breeds were identified using pairwise comparisons based on F_{ST} , which quantifies differences in allele frequencies among populations [37]. F_{ST} values range from 0, indicating little or no population differentiation, to 1, indicating complete fixation of alternative alleles.

SNPs within the upper 0.1% of the empirical F_{ST} distribution were classified as putative differentiation outliers. This threshold was applied independently to each comparison using the same dataset of 30,839 SNPs. Because this percentile-based threshold represents an empirical heuristic rather than a formal inferential test, the

identified loci were considered exploratory candidate markers requiring independent validation.

Search for loci under selection pressure

Candidate loci under putative selection pressure were investigated using three complementary approaches: ROH islands, pairwise F_{ST} analysis, and cross-population extended haplotype homozygosity (XP-EHH).

ROH were identified using the consecutiveRUNS method implemented in the R package detectRUNS version 0.9.6 (<https://cran.r-project.org/web/packages/detectRUNS/index.html>) [39]. To minimize underestimation of longer ROH, one missing genotype and one heterozygous SNP were permitted within a run. The minimum ROH length was set at 500 kb to reduce inclusion of short homozygous segments commonly distributed throughout the genome.

The minimum number of SNPs required to define an ROH was calculated using the following equation:

$$l = \frac{\log_e \frac{\alpha}{n_s \cdot n_i}}{\log_e (1 - \overline{het})}$$

where n_s is the number of SNPs genotyped per individual, n_i is the number of genotyped individuals, α represents the assumed proportion of false-positive ROH (0.05), and the mean heterozygosity across all SNPs was incorporated into the calculation. The resulting minimum was 18 SNPs per ROH.

ROH islands were defined as genomic regions shared by at least 25% of animals within a population. This threshold was selected to allow detection of recurrent homozygous regions in the comparatively less inbred Kalmyk population while avoiding excessive dependence on the higher homozygosity levels observed in commercial beef cattle. The average genomic inbreeding coefficient was $F_{ROH} = 0.050 \pm 0.001$ in KALMYK compared with $F_{ROH} = 0.133 \pm 0.007$ in specialized beef breeds.

The 25% threshold was also supported by the observed distribution of the detected ROH islands. The *Bos taurus* autosome (BTA)5 region at 75.35–77.00 Mb occurred in the Kalmyk, beef, and steppe groups despite differences in their inbreeding levels. Two additional Kalmyk-specific regions were identified on BTA7 (56.71–57.15 Mb) and BTA13 (53.57–54.36 Mb). All ROH coordinates were reported relative to the ARS-UCD1.2 assembly.

Crucially, the minimum length (500 kb) and minimum SNP count (18) criteria were strictly applied during the initial detection of individual ROH segments within each animal. In contrast, a population level ROH island was defined as a genomic region where these individual ROH segments overlapped in at least 25% of the animals within a group. Because the boundaries of a population level island are determined by the outer edges of the overlapping individual segments, the physical length and the number of SNPs within the final reported ROH island may be shorter or fewer than the thresholds applied to individual ROH detection.

For F_{ST} -based identification of candidate genomic regions, a 400-kb window extending ± 200 kb from each outlier SNP was considered. SNPs within the upper 0.1% of the empirical F_{ST} distribution were used to define regions showing the strongest population differentiation. When multiple outlier SNPs were located in close physical proximity, their respective 400-kb windows were merged into a single, continuous candidate genomic region to prevent the artificial fragmentation of broad selection signals. Because F_{ST} does not intrinsically provide inferential p-values without specification of a null model, these signals were treated as hypothesis-generating candidates rather than confirmatory evidence of selection.

XP-EHH analysis was performed to identify extended haplotype differentiation between populations [40]. In each pairwise comparison, Kalmyk cattle were designated as the test population, whereas Hereford, KZWH, Red Steppe, or MONG served as the corresponding reference population. XP-EHH was calculated using the R package rehh version 3.2.2 [41].

XP-EHH was implemented in unpolarized mode because reliable ancestral-state information was unavailable for BovineSNP50 markers in Turano-MONG. Scores were standardized using the built-in rehh procedure under the assumption of a standard normal distribution. The parameters $\text{min.maf} = 0.05$, $\text{min.snp} = 5$, and $\text{include.sex} = \text{FALSE}$ were applied. Multiple testing across all 30,839 SNPs was controlled using the Benjamini–Hochberg false discovery rate (FDR) procedure, applied separately for each pairwise comparison. SNPs with FDR-adjusted q-values < 0.05 were considered statistically supported signals. In these comparisons, Kalmyk cattle were designated as the test (numerator) population; therefore, positive XP-EHH values indicate longer haplotypes (putative selection) in the Kalmyk population, whereas negative values indicate longer haplotypes in the reference population.

To identify contiguous candidate regions, significant SNPs (FDR $q < 0.05$) were clustered if they were located

within a maximum physical distance of 500 kb from each other. The boundaries of each clustered region were defined by the outermost significant SNPs within the cluster, extended by a 500-kb flanking window and downstream. Isolated significant SNPs that did not form part of a contiguous spatial cluster were retained as primary data, but were not defined as candidate regions.

The F_{ST} and XP-EHH analyses were performed using two principal comparison frameworks: Kalmyk cattle versus a combined specialized beef group comprising Hereford and KZWH cattle, and Kalmyk cattle versus a combined steppe group comprising Red Steppe and MONG.

Overall pipeline

All genomic and statistical analyses were conducted using PLINK version 1.9 [28] and R version 4.2.2 (R Core Team, 2022). The R packages used included *diversity* version 1.9.90, *ggplot2* version 3.5.0, *BITE* version 2.2.1, *detectRUNS* version 0.9.6, and *rehh* version 3.2.2. *SplitsTree* version 4.14.5 was used for phylogenetic-network reconstruction.

The three Illumina platforms—BovineSNP50 v1 (53,599 SNPs), BovineSNP50 v3 (53,218 SNPs), and BovineHD BeadChip (777,962 SNPs)—were intersected to retain 44,774 common SNPs. Quality control filtering removed 13,935 SNPs and resulted in a final dataset comprising 30,839 SNPs across 312 animals. Where LD pruning was required, an additional 2907 SNPs were removed, producing a dataset of 27,932 SNPs. The full 30,839-SNP dataset was retained for the selection-signature analyses where LD pruning was not required (Figure 2).

No animals were excluded at the sample-level call rate threshold, and cryptic relatedness analysis identified no pairs requiring removal. Although the Hereford group contained 27 animals, all individuals passed the specified quality control criteria, with sample call rates ≥ 0.961 .

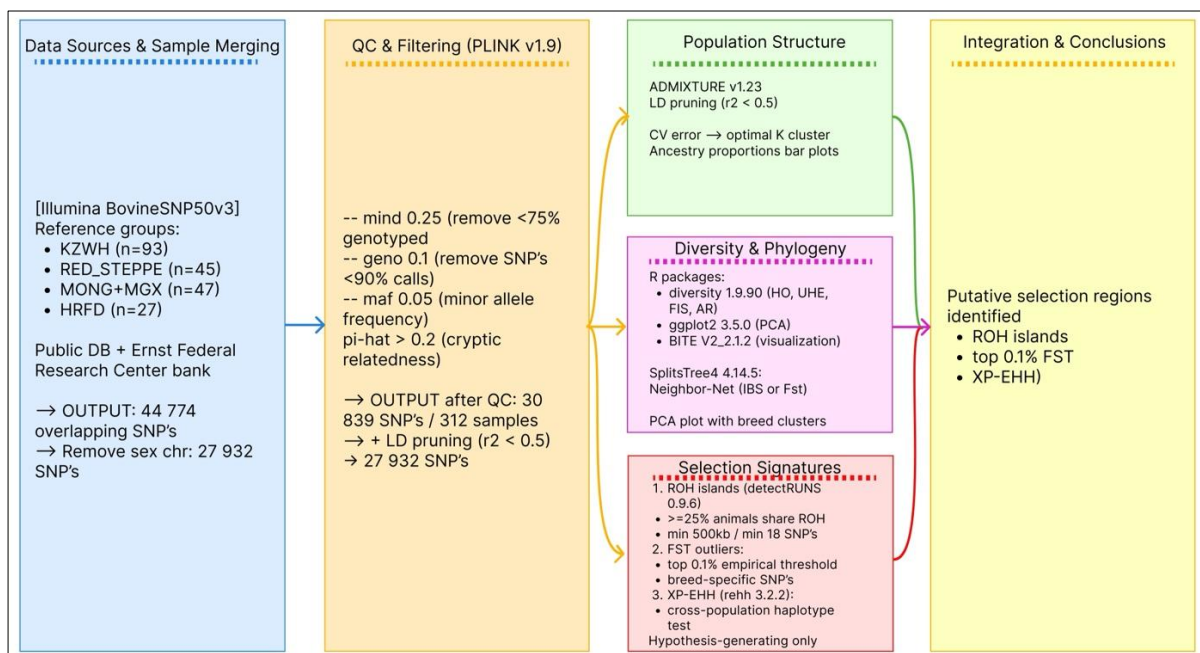


Figure 2: Overall bioinformatic pipeline used for quality control, population structure analysis, genetic diversity assessment, and identification of candidate selection signatures in Kalmyk cattle.

Annotation of genes

Structural and functional annotation of genes located within candidate genomic regions was performed using Ensembl (www.ensembl.org) release 112 and DAVID version 2025 (<https://davidbioinformatics.nih.gov/>), both accessed on January 1, 2026. The *Bos taurus* ARS-UCD1.2 genome assembly was used as the background gene set for DAVID analyses. The Cattle QTLdb release 59 (<https://www.animalgenome.org/>), accessed on April 20, 2026, was additionally consulted to compare identified genomic regions with previously reported quantitative trait loci.

A DAVID cluster enrichment score ≥ 1.3 was used as the threshold for enrichment interpretation according to the analytical criterion specified in the study.

Graphical representation of study design

The free version of Jay Flow was used to generate illustrative cattle images of Mongolian, Red Steppe,

Hereford, and Kazakh White-Headed cattle for the purpose of graphical representation of study design. Prompts describing typical breed characteristics included coat color, head coloration, average body weight, horn morphology and position, height at the withers, and rump characteristics.

RESULTS

Analysis of population-genetic and phylogenetic relationships of KALMYK

After all quality control and filtering steps, the final dataset comprised 312 animals (KALMYK, $n = 100$; HRFD, $n = 27$; KZWH, $n = 93$; RED_STEPPE, $n = 45$; and MONG, $n = 47$) and 30,839 autosomal SNPs for selection-signature analyses. An additional LD-pruned subset of 27,932 SNPs was used for population structure analyses, including PCA, ADMIXTURE, and IBS-based network reconstruction. No animals were excluded during quality control filtering.

Genetic diversity among the studied cattle populations was assessed using A_R , observed and unbiased expected heterozygosity, and the unbiased inbreeding coefficient (Table 1).

Table 1: Genetic diversity parameters of the studied cattle populations based on SNP markers.

Population	n	H_O (M $\pm$ SE)	uH_E (M $\pm$ SE)	A_R (M $\pm$ SE)	uF_IS (95% CI)
KALMYK	100	0.356 $\pm$ 0.001	0.355 $\pm$ 0.001	1.890 $\pm$ 0.002	-0.001 (-0.002–0.000)
HRFD	27	0.321 $\pm$ 0.001	0.346 $\pm$ 0.001	1.859 $\pm$ 0.002	0.069 (0.067–0.071)
KZWH	93	0.362 $\pm$ 0.001	0.359 $\pm$ 0.001	1.891 $\pm$ 0.002	-0.009 (-0.010–0.008)
MONG	47	0.345 $\pm$ 0.001	0.352 $\pm$ 0.001	1.891 $\pm$ 0.001	0.019 (0.017–0.021)
RED_STEPPE	45	0.347 $\pm$ 0.001	0.361 $\pm$ 0.001	1.892 $\pm$ 0.002	0.038 (0.036–0.040)

A_R = Allelic richness, CI = Confidence interval, H_O = Observed heterozygosity, HRFD = Hereford, KALMYK = Kalmyk, KZWH = Kazakh White-Headed, M = Mean, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, SE = Standard error, SNP = Single-nucleotide polymorphism, uF_IS = Unbiased inbreeding coefficient, uH_E = Unbiased expected heterozygosity.

The Kalmyk population showed an H_O of 0.356 ± 0.001 and unbiased expected heterozygosity of 0.355 ± 0.001 . The uF_IS estimate was -0.001 (95% CI = -0.002–0.000), indicating little evidence of heterozygote deficit or excess in the analyzed population. A_R in KALMYK was 1.890 ± 0.002 . Among the studied populations, the lowest A_R was observed in HRFD (1.859 ± 0.002).

IBS-based genetic distances were used to construct a Neighbor-Net network to visualize clustering among individuals from the studied populations (Figure 3).

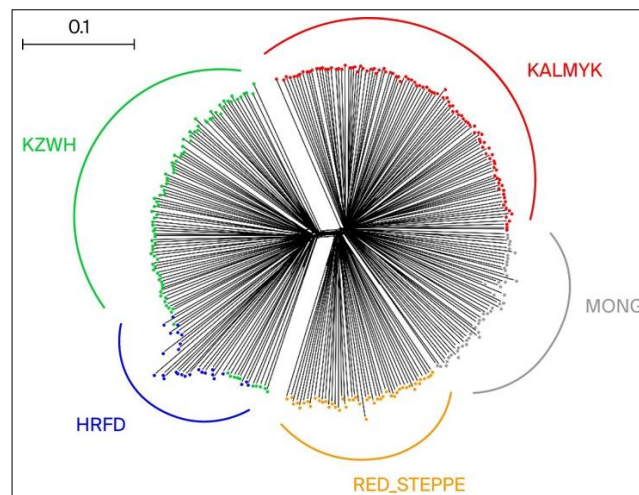


Figure 3: Neighbor-Net network based on IBS genetic distances calculated using 27,932 SNPs. HRFD = Hereford, KALMYK = Kalmyk, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, IBS = Identical-by-state, SNP = Single-nucleotide polymorphism.

Interbreed genetic differentiation was further evaluated using pairwise F_{ST} estimates (Table 2).

Table 2: Pairwise F_{ST} estimates among the studied cattle populations based on 27,932 SNPs.

Breed	HRFD	KALMYK	KZWH	MONG	RED_STEPPE
HRFD	–				
KALMYK	0.113	–			
KZWH	0.028	0.069	–		

Breed	HRFD	KALMYK	KZWH	MONG	RED_STEPPE
MONG	0.121	0.019	0.076	–	
RED_STEPPE	0.099	0.037	0.058	0.043	–

HRFD = Hereford, KALMYK = Kalmyk, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, SNP = Single-nucleotide polymorphism.

Pairwise F_{ST} relationships were additionally visualized using a Neighbor-Net network (Figure 4). The smallest F_{ST} values were observed between KALMYK and MONG ($F_{ST} = 0.019$) and between HRFD and KZWH ($F_{ST} = 0.028$), indicating comparatively lower genetic differentiation within these population pairs.

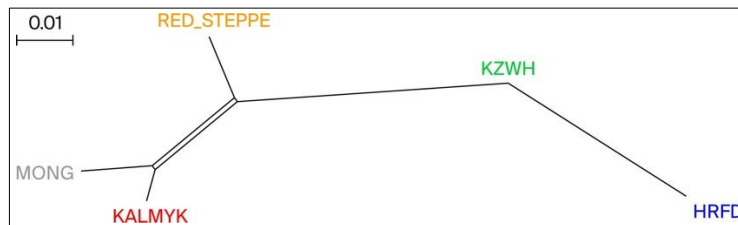


Figure 4: Neighbor-Net network based on pairwise F_{ST} estimates among the studied cattle populations.

Population structure was further evaluated using ADMIXTURE analysis (Figure 5). The optimal number of ancestral clusters was $K = 5$, corresponding to the lowest cross-validation error across K values from 1 to 10.

At $K = 2$, Hereford and KZWH cattle shared a predominant ancestry component that differentiated them from the remaining populations. At $K = 3$, Red Steppe cattle formed a more distinct component, while KALMYK and MONG retained shared ancestry components. At $K = 4$, MONG formed a more distinct cluster, and shared ancestry components remained evident between Kalmyk and Red Steppe cattle. At the optimal $K = 5$, KZWH cattle became further differentiated from HRFD, whereas KALMYK and MONG continued to exhibit shared ancestry components.

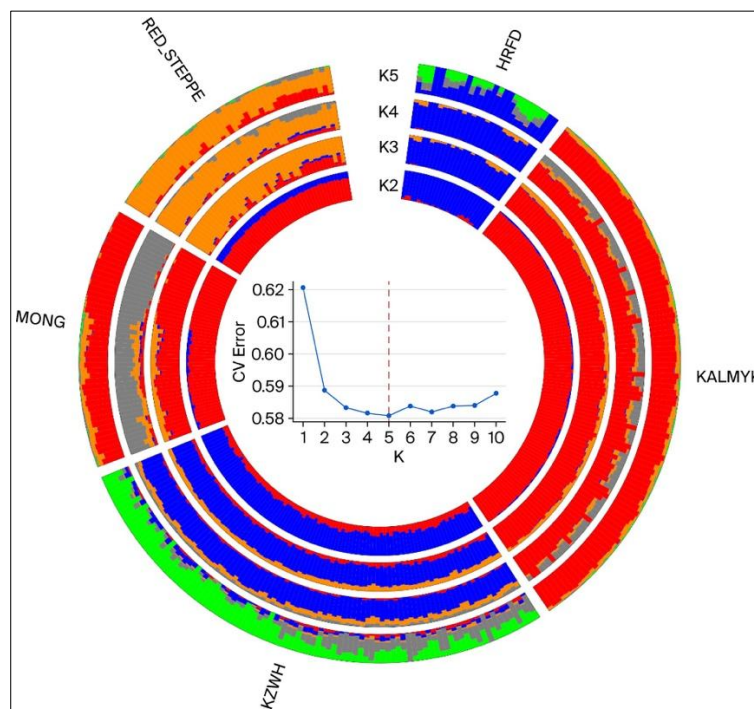


Figure 5: Population structure of the studied cattle populations inferred using ADMIXTURE. The dashed line indicates the optimal number of ancestral clusters ($K = 5$) determined from the lowest cross-validation error. Individual ancestry proportions are represented by different colors.

Identification of differentiating SNPs distinguishing KALMYK from other studied breeds

Candidate SNPs differentiating KALMYK from the other studied breeds were identified using F_{ST} -based comparisons of allele frequencies. Based on the population-genetic and phylogenetic results, two comparison groups were constructed: KALMYK versus the combined specialized beef group comprising Hereford and KZWH, and KALMYK versus the combined steppe group comprising Red Steppe and MONG.

These comparisons were used to identify loci showing the strongest differentiation between KALMYK and

the specialized beef group (Figure 6A) and between KALMYK and the steppe group (Figure 6B).

The mean F_{ST} value for the comparison between KALMYK and the specialized beef group was 0.0696, whereas the corresponding mean for the comparison with the steppe group was 0.0158. Thus, genome-wide differentiation was greater between KALMYK and the specialized beef group than between KALMYK and the steppe group.

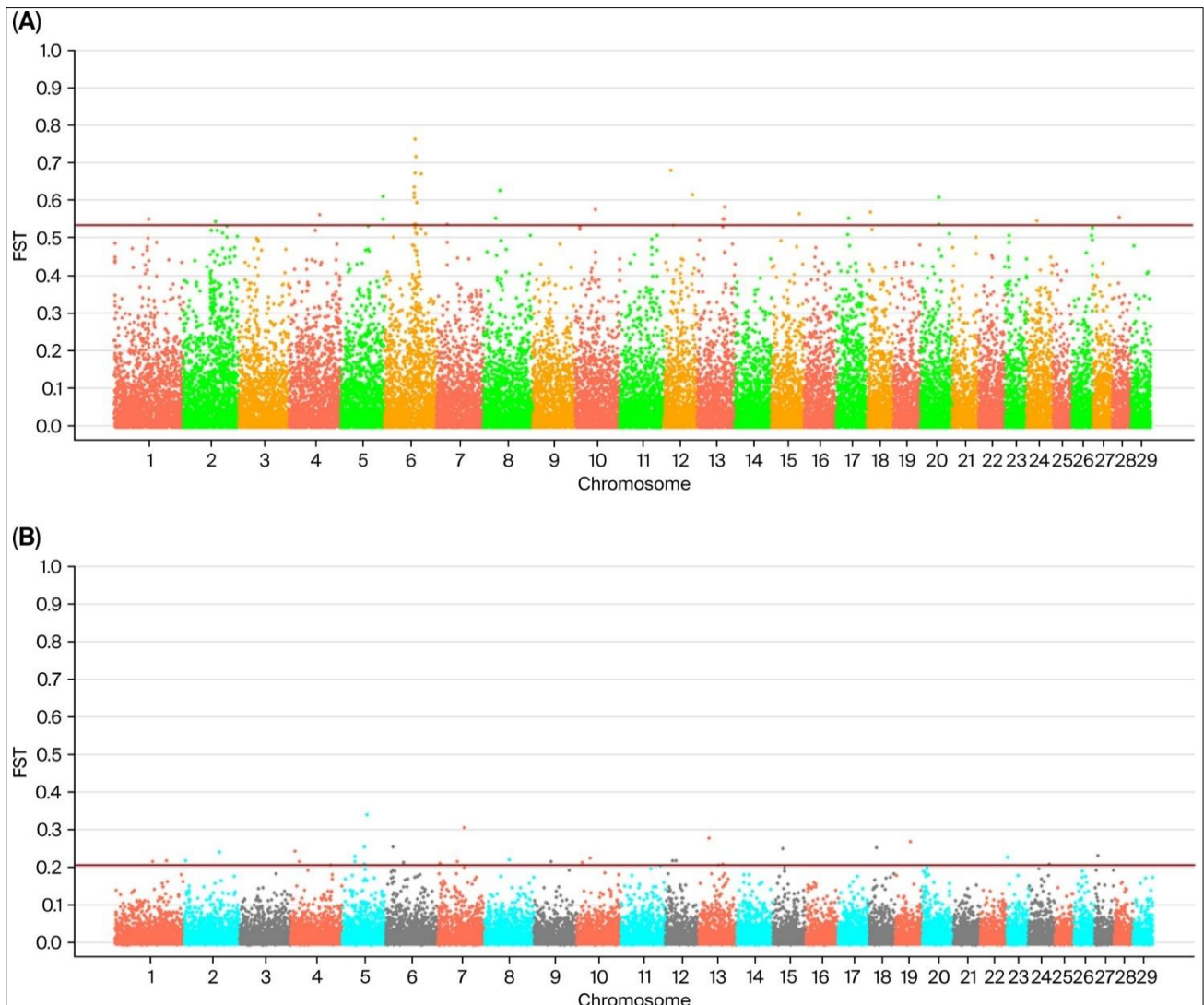


Figure 6: Distribution of F_{ST} values across bovine autosomes. (A) Kalmyk cattle versus the combined Hereford and Kazakh White-Headed group (KALM vs. HRFD+KZWH). (B) Kalmyk cattle versus the combined Red Steppe and Mongolian group (KALM vs. RED_STEPPE+MONG). The x-axis represents bovine autosomes according to chromosomal position, and the y-axis represents F_{ST} values. The horizontal line indicates the upper 0.1% empirical F_{ST} threshold. Colors distinguish chromosomes and have no additional biological meaning.

For each comparison, 31 SNPs exceeded the upper 0.1% empirical F_{ST} threshold (Table 3).

Table 3: Chromosomal distribution of candidate differentiating SNPs identified in comparisons of Kalmyk cattle with specialized beef and steppe cattle groups.

Chromosome	KALM vs. HRFD+KZWH, n	Percentage	KALM vs. RED_STEPPE+MONG, n	Percentage
1	1	3.23	2	6.45
2	1	3.23	2	6.45
4	1	3.23	3	9.68
5	2	6.45	5	16.13
6	10	32.26	2	6.45
7	1	3.23	3	9.68
8	2	6.45	1	3.23
9	—	—	1	3.23
10	1	3.23	2	6.45

Chromosome	KALM vs. HRFD+KZWH, n	Percentage	KALM vs. RED_STEPPE+MONG, n	Percentage
12	2	6.45	2	6.45
13	3	9.68	2	6.45
15	1	3.23	1	3.23
17	1	3.23	—	—
18	1	3.23	1	3.23
19	—	—	1	3.23
20	2	6.45	—	—
23	—	—	1	3.23
24	1	3.23	1	3.23
27	—	—	1	3.23
28	1	3.23	—	—
Total	31	100	31	100

HRFD = Hereford, KALM = Kalmyk, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, SNP = Single-nucleotide polymorphism.

In the comparison with specialized beef cattle, candidate differentiating SNPs were distributed across 16 chromosomes, with the largest proportions located on BTA6 (32.26%) and BTA13 (9.68%) (Figure 6A). In the comparison with the steppe group, candidate SNPs were distributed across 17 chromosomes, with the highest proportion on BTA5 (16.13%), followed by BTA4 and BTA7 (9.68% each) (Figure 6B).

The chromosomal positions of the candidate SNPs identified in both comparisons are presented in Table 4.

Table 4: Candidate SNPs differentiating Kalmyk cattle from specialized beef and steppe cattle groups.

BTA	KALM vs. HRFD+KZWH, SNP (position, bp)	KALM vs. RED_STEPPE+MONG, SNP (position, bp)
1	Hapmap41250-BTA-38429 (79,548,115)	Hapmap55404-rs29024815 (84,680,569); Hapmap42953-BTA-48231 (117,593,495)
2	BTA-96609-no-rs (75,615,637)	ARS-BFGL-NGS-55090 (1,343,910); ARS-BFGL-NGS-71649 (82,271,502)
4	BTA-116026-no-rs (70,045,272)	ARS-BFGL-NGS-74263 (11,344,256); BTA-72234-no-rs (20,924,989); Hapmap47201-BTA-121648 (91,525,544)
5	ARS-BFGL-NGS-56171 (114,768,420); ARS-BFGL-NGS-1069 (114,829,438)	Hapmap42150-BTA-25138 (32,649,736); ARS-BFGL-NGS-717 (32,692,339); Hapmap51479-BTA-66720 (63,990,439); Hapmap55537-rs29016129 (64,036,142); Hapmap28985-BTA-73836 (69,996,344)
6	Hapmap40668-BTA-120410 (67,216,723); ARS-BFGL-NGS-20354 (67,835,794); Hapmap38628-BTA-76690 (67,991,798); Hapmap60836-rs29027147 (68,560,206); Hapmap38134-BTA-107929 (68,700,887); BTA-76626-no-rs (69,310,878); ARS-BFGL-NGS-7423 (70,673,035); ARS-BFGL-NGS-117025 (71,951,614); BTA-76842-no-rs (73,643,517); BTA-64031-no-rs (88,577,622)	ARS-BFGL-NGS-40548 (16,783,120); BTB-00251059 (40,731,922)
7	BTB-01074080 (26,707,302)	ARS-BFGL-NGS-3205 (3,821,837); BTB-00309643 (45,629,869); Hapmap54307-rs29014078 (61,965,719)
8	BTB-00258020 (32,364,528); ARS-BFGL-NGS-83850 (43,348,930)	ARS-BFGL-NGS-23181 (58,523,210)
9	—	Hapmap24598-BTA-83620 (44,603,627)
10	ARS-BFGL-NGS-111546 (49,422,403)	ARS-BFGL-NGS-10428 (13,860,736); Hapmap44110-BTA-63281 (32,833,799)
12	Hapmap30016-BTA-149504 (20,192,311) ¹ ; Hapmap48930-BTA-87766 (77,636,724)	Hapmap30016-BTA-149504 (20,192,311) ¹ ; BTA-83050-no-rs (29,537,962)
13	ARS-BFGL-NGS-24352 (59,194,693); ARS-BFGL-NGS-17925 (63,328,898); UA-IFASA-7527 (63,399,473)	ARS-BFGL-NGS-113296 (24,721,448); ARS-BFGL-NGS-30853 (57,059,856)
15	BTB-00613489 (70,621,669)	BTA-36195-no-rs (27,095,061)
17	Hapmap55008-rs29022004 (29,808,252)	—
18	ARS-BFGL-NGS-1162 (8,292,883)	Hapmap47005-BTA-42576 (18,659,338)
19	—	ARS-BFGL-NGS-6633 (36,585,153)
20	Hapmap42236-BTA-50618 (42,133,960); Hapmap38112-BTA-50631 (42,543,356)	—

BTA	KALM vs. HRFD+KZWH, SNP (position, bp)	KALM vs. RED_STEPPE+MONG, SNP (position, bp)
23	–	Hapmap44605-BTA-55952 (3,049,312)
24	ARS-BFGL-NGS-63047 (25,666,802)	ARS-BFGL-NGS-39670 (47,080,160)
27	–	ARS-BFGL-NGS-21780 (11,954,777)
28	ARS-BFGL-NGS-18928 (17,776,649)	–

BTA = *Bos taurus* autosome, HRFD = Hereford, KALM = Kalmyk, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, SNP = Single-nucleotide polymorphism. SNP identifiers follow the Illumina DNA-chip manifest; genomic positions are reported according to the *Bos taurus* ARS-UCD1.2 assembly (GCF_002263795.3). ¹SNP shared between both comparison groups.

A total of 31 candidate SNPs exceeded the empirical F_{ST} threshold in each comparison, producing 62 comparison-specific observations. One locus, Hapmap30016-BTA-149504 on BTA12 (20,192,311 bp), was shared between both comparisons. After accounting for this shared locus, 61 unique candidate SNPs remained: 30 unique to the specialized beef comparison, 30 unique to the steppe comparison, and one common to both groups.

Among the 61 candidate SNPs, 30 were located in intronic regions, 24 in intergenic regions, and seven in regions upstream or downstream of annotated gene variants (Table S1). No nonsynonymous coding variants were identified among these loci. Accordingly, the identified SNPs represent candidate differentiating markers rather than protein-altering variants.

Technical characteristics of the 61 candidate SNPs were also evaluated in the combined dataset of 312 animals. The mean minor allele frequency was 0.36, with values ranging from 0.10 to 0.40, and the mean call rate was 0.97, ranging from 0.90 to 1.00. All candidate SNPs met the predefined quality control criteria for minor allele frequency and call rate. SNP-specific quality control statistics are presented in Table S2.

Search for loci under selection pressure in the genome of KALMYK

Candidate genomic regions potentially influenced by selection were investigated using three complementary approaches: ROH islands, F_{ST} -based differentiation, and XP-EHH analysis.

Genomic coordinates were reported relative to the *Bos taurus* ARS-UCD1.2 assembly to maintain consistency with current gene annotation and downstream functional analyses.

Comparison of the F_{ST} and XP-EHH results identified a convergent candidate region on BTA6 spanning approximately 67.45–73.64 Mb in the comparison between KALMYK and the combined Hereford and KZWH group. This region contained eight F_{ST} outlier SNPs and 25 candidate genes. No other overlapping regions between the F_{ST} and XP-EHH approaches were identified in the supplied analysis.

Analysis of F_{ST} values in pairwise comparisons of breeds

Genomic regions extending 400 kb around candidate SNPs (± 200 kb from the SNP position) were considered candidate regions under putative selection. Candidate SNPs were defined as those within the upper 0.1% of the empirical F_{ST} distribution. Structural annotation of the identified genomic regions was subsequently performed (Tables 5 and 6). Across the two comparison groups, 202 genes located on 20 autosomes were identified within these candidate regions.

Table 5: Candidate genomic regions identified by F_{ST} analysis in the studied cattle groups.

KALM vs. HRFD+KZWH	KALM vs. RED_STEPPE+MONG
1:79350408–79747985	1:84480578–84880446
–	1:117393857–117792612
2:75417085–75815532	2:1144037–1543274
–	2:82072292–82471650
4:69845354–70258554	4:11144870–11531949
–	4:20729331–21126942
–	4:91325556–91725518
5:114404591–115016944	5:32451230–32892601
–	5:63790828–64240956
–	5:69798385–70194783
6:67019087–73843581	6:16677031–16983098
6:88377608–88777639	6:40532328–40931188
7:26507146–26910962	7:3621877–4015817
–	7:45434753–45830843
–	7:61765932–62165912
8:32164615–32565061	8:58323064–58722242

KALM vs. HRFD+KZWH	KALM vs. RED_STEPPE+MONG
8:43145488–43554505	–
–	9:44476887–44790766
10:49222432–49622493	10:13660748–14051034
–	10:32634231–33033065
12:19968336–20390673 ¹	12:19968336–20390673 ¹
12:77436359–77836604	12:29337895–29737957
13:58996324–59394891	13:24523670–24921658
13:63131933–63604652	13:56858803–57261592
15:70422380–70820864	15:26895722–27295776
17:29607505–30008075	–
18:8092995–8478226	18:18460550–18856802
–	19:36387501–36785017
20:41934128–42743462	–
–	23:2850014–3248688
24:25466981–25865504	24:46878945–47280309
–	27:11755119–12154752
28:17576544–17976738	–

Coordinates are presented as BTA:start–end, where BTA = *Bos taurus* autosome, start = Start position of the genomic region, and end = End position of the genomic region. HRFD = Hereford, KALM = Kalmyk, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe. All genomic coordinates are reported relative to the ARS-UCD1.2 assembly (GCF_002263795.3).

¹Region shared between both comparisons.

Table 6: Chromosomal distribution of genes identified within candidate genomic regions based on F_{ST} analysis.

BTA	KALM vs. HRFD+KZWH, n	Percentage	KALM vs. RED_STEPPE+MONG, n	Percentage
1	3	2.97	6	5.94
2	–	–	5	4.95
4	2	1.98	7	6.93
5	5	4.95	16	15.84
6	47	46.53	1	0.99
7	3	2.97	33	32.67
8	5	4.95	–	–
9	–	–	2	1.98
10	1	0.99	2	1.98
12	2	1.98	3	2.97
13	9	8.91	4	3.96
15	–	–	1	0.99
17	9	8.91	–	–
18	5	4.95	7	6.93
19	–	–	9	8.91
20	3	2.97	–	–
23	–	–	4	3.96
24	5	4.95	1	0.99
28	2	1.98	–	–

BTA = *Bos taurus* autosome, HRFD = Hereford, KALM = Kalmyk, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe.

In the comparison between KALMYK and the combined specialized beef group comprising Hereford and KZWH, the largest proportions of genes within candidate regions were located on BTA6 (46.53%), BTA13 (8.91%), and BTA17 (8.91%). In the comparison between KALMYK and the combined steppe group comprising Red Steppe and MONG, the largest proportions were located on BTA7 (32.67%), BTA5 (15.84%), and BTA19 (8.91%).

Genes located within the candidate genomic regions identified in the two pairwise comparisons are presented in Table 7.

Table 7: Genes identified within candidate genomic regions differentiating Kalmyk cattle from the comparison groups.

Genomic region	Comparison group	Genes localized within the identified region
1:79350408–79747985	HRFD+KZWH	<i>BCL6, RTP2, SST</i>
1:117393857–117792612	RED_STEPPE+MONG	<i>ERICH6, SELT, EIF2A, SNORA42, SERP1, TSC22D2</i>
2:1144037–1543274	RED_STEPPE+MONG	<i>TUBGCP5, CCDC115, IMP4, PTPN18, AMER3</i>

Genomic region	Comparison group	Genes localized within the identified region
4:11144870–11531949	RED_STEPPE+MONG	<i>TFPI2, GNGT1, GNG11, BET1</i>
4:20729331–21126942	RED_STEPPE+MONG	<i>ARL4A</i>
4:69845354–70258554	HRFD+KZWH	<i>bta-mir-148a, U6</i>
4:91325556–91725518	RED_STEPPE+MONG	<i>ZNF800, 7SK</i>
5:32451230–32849889	RED_STEPPE+MONG	<i>HDAC7, SLC48A1, RAPGEF3, ENDOU, RPAP3</i>
5:32493513–32892601	RED_STEPPE+MONG	<i>SLC48A1, RAPGEF3, ENDOU, RPAP3</i>
5:63790828–64190810	RED_STEPPE+MONG	<i>UHRF1BP1L, ACTR6</i>
5:63837620–64240956	RED_STEPPE+MONG	<i>UHRF1BP1L, ACTR6, SCYL2</i>
5:69798385–70194783	RED_STEPPE+MONG	<i>RFX4, RIC8B</i>
5:114404591–114959387	HRFD+KZWH	<i>PARVB, MGC148449</i>
5:114342554–115016944	HRFD+KZWH	<i>PARVB, MGC148449, KIAA1644</i>
6:40532328–40931188	RED_STEPPE+MONG	<i>KCNIP4</i>
6:67019087–67416609	HRFD+KZWH	<i>SLAIN2, SLC10A4, ZAR1, FRYL, 5S_rRNA</i>
6:67634038–68034750	HRFD+KZWH	<i>DCUN1D4, LRRC66, SGCB, SPATA18</i>
6:67791080–68192058	HRFD+KZWH	<i>LRRC66, SGCB, SPATA18</i>
6:68360587–68759387	HRFD+KZWH	<i>bta-mir-4449, SNORA26, RASL11B, SCFD2</i>
6:68500855–68900547	HRFD+KZWH	<i>RASL11B, SCFD2</i>
6:69110163–69510878	HRFD+KZWH	<i>CLOCK, PDCL2, NMU</i>
6:70474595–70838728	HRFD+KZWH	<i>U6, KDR, SRD5A3, RESTB, NOA1A, POLR2B, IGFBP7</i>
6:71751570–72151827	HRFD+KZWH	<i>PPAT, PAICS, U6, SRP72, ARL9, HOP, U6atac, SPINK2</i>
6:73443602–73843581	HRFD+KZWH	<i>Y_rRNA</i>
6:88377608–88777639	HRFD+KZWH	<i>ALB, AFP, AFM, RASSF6, 7SK, RCHY1, THAP6, ODAPH, CDKL2, G3BP2</i>
7:3621877–4015817	RED_STEPPE+MONG	<i>ATP13A1, GMIP, LPAR2, PBX4, U6, CILP2, YJEFN3, NDUFA13, TSSK6, GATAD2A, MAU2, SF4, TM6SF2, HAPLN4</i>
7:26507146–26910962	HRFD+KZWH	<i>MEGF10, C5orf63, U2</i>
7:45434753–45830843	RED_STEPPE+MONG	<i>C7H5ORF15, VDAC1, TCF7, SKP1, NEUROG1, CXCL14, SLC25A48, IL9, FBXL21, LECT2</i>
7:61765932–62165912	RED_STEPPE+MONG	<i>NDST1, bta-mir-2461, SYNPO, MYOZ3, 5S_rRNA, RBM22, DCTN4, U2, MGC128424</i>
8:43145488–43554505	HRFD+KZWH	<i>DMRT2, 5S_rRNA, DMRT3, DMRT1, KANK1</i>
9:44476887–44790766	RED_STEPPE+MONG	<i>PREP, HACE1</i>
10:13660748–14051034	RED_STEPPE+MONG	<i>SMAD3, MGC152585</i>
10:49222432–49622493	HRFD+KZWH	<i>RORA</i>
12:19968336–20390673 ¹	HRFD+KZWH	<i>RNASEH2B</i>
12:19968336–20390673 ¹	RED_STEPPE+MONG	<i>RNASEH2B</i>
12:29337895–29737957	RED_STEPPE+MONG	<i>bta-mir-2299, B3GALT1</i>
12:77436359–77836604	HRFD+KZWH	<i>NALCN</i>
13:24523670–24921658	RED_STEPPE+MONG	<i>7SK, SNORA18</i>
13:56858803–57261592	RED_STEPPE+MONG	<i>EDN3, ZNF831</i>
13:58996324–59394891	HRFD+KZWH	<i>TFAP2C</i>
13:63131933–63528733	HRFD+KZWH	<i>PXMP4, ZNF341, CHMP4B, RALY, EIF2S2</i>
13:63202511–63604652	HRFD+KZWH	<i>CHMP4B, RALY, EIF2S2</i>
15:26895722–27295776	RED_STEPPE+MONG	<i>snoZ6</i>
17:29607505–30008075	HRFD+KZWH	<i>U2, MGC165715, SNORA42, MFSD8, PLK4, HSPA4L, SLC25A31, INTU, U6</i>
18:8092995–8478226	HRFD+KZWH	<i>bta-mir-2883, CMIP, PLCG2, 5S_rRNA, SNORA70</i>
18:18460550–18856802	RED_STEPPE+MONG	<i>HEATR3, snoU13, snoU2_19, snoU2-30, PAPD5, ADCY7, BRD7</i>
19:36387501–36785017	RED_STEPPE+MONG	<i>TMEM92, U6, COL1A1, SGCA, SAMD14, PDK2, ITGA3, DLX3, DLX4</i>
20:41934128–42334623	HRFD+KZWH	<i>C20H5ORF22, DROSHA, CDH6</i>
23:2850014–3248688	RED_STEPPE+MONG	<i>RAB23, BAG2, ZNF451, BEND6</i>
24:25466981–25865504	HRFD+KZWH	<i>TTR, DSG2, DSG3, DSG4, DSG1</i>
24:46878945–47280309	RED_STEPPE+MONG	<i>SKOR2</i>
28:17576544–17976738	HRFD+KZWH	<i>C10ORF107, ARID5B</i>

Coordinates are presented as BTA:start–end. HRFD+KZWH = Combined specialized beef group comprising Hereford and Kazakh White-Headed cattle, RED_STEPPE+MONG = Combined steppe group comprising Red Steppe and Mongolian cattle. All genomic coordinates are reported relative to the ARS-UCD1.2 assembly (GCF_002263795.3).

¹Genomic region shared between both comparisons.

One candidate region on BTA12 (19,968,336–20,390,673 bp) was identified in both comparisons. This region contained *RNASEH2B*. The shared F_ST outlier SNP Hapmap30016-BTA-149504, located at 20,192,311 bp, was

situated within an intronic region of *RNASEH2B*. Thus, this locus represented the only F_{ST} candidate region shared between the specialized beef and steppe comparisons.

Functional annotation using DAVID was performed for genes within regions specific to each comparison. For the comparison between KALMYK and the specialized beef group, 12 functional annotation clusters were generated, of which three met the predefined enrichment score threshold of ≥ 1.3 . These clusters included genes associated with sexual differentiation, cell adhesion, cellular morphogenesis, cytoskeletal development, and protein-related processes.

For the comparison between KALMYK and the steppe group, 10 functional annotation clusters were generated, of which one met the predefined enrichment score threshold of ≥ 1.3 . The cluster included genes associated with signaling pathways related to stem-cell pluripotency, synaptic processes, myopathy-related functions, and circadian regulation.

Search for regions of excessive homozygosity (ROH islands)

Analysis of homozygosity patterns revealed differences in genomic inbreeding among the studied cattle groups (Table 8). The combined specialized beef group (HRFD+KZWH) had the highest F_{ROH} (0.133 ± 0.007), compared with 0.050 ± 0.001 in KALMYK and 0.068 ± 0.004 in the combined steppe group. The HRFD+KZWH group also had the greatest mean total ROH length and mean number of ROH segments.

Table 8: Homozygosity and genomic inbreeding parameters in the studied cattle groups.

Group	n	ROH length (M $\pm$ SE)	ROH number (M $\pm$ SE)	F _{ROH} (M $\pm$ SE)
HRFD+KZWH	120	332.46 $\pm$ 17.91	102.56 $\pm$ 1.89	0.133 $\pm$ 0.007
KALMYK	100	124.78 $\pm$ 3.62	61.14 $\pm$ 0.83	0.050 $\pm$ 0.001
RED_STEPPE+MONG	95	169.01 $\pm$ 9.53	68.11 $\pm$ 2.23	0.068 $\pm$ 0.004

F_{ROH} = Genomic inbreeding coefficient based on runs of homozygosity, HRFD = Hereford, KALMYK = Kalmyk, KZWH = Kazakh White-Headed, M = Mean, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, ROH = Runs of homozygosity, SE = Standard error.

Genomic regions characterized by recurrent ROH were subsequently identified in each cattle group (Table 9).

Table 9: ROH islands identified in the studied cattle groups.

Group	BTA	Beginning (bp)	End (bp)	Length (Mb)	Number of SNPs
HRFD+KZWH	1	1,566,539	3,492,221	1.93	36
HRFD+KZWH	1	70,656,139	71,132,759	0.48	6
HRFD+KZWH	1	73,178,671	73,971,461	0.79	10
HRFD+KZWH	1	76,119,473	76,627,997	0.51	8
HRFD+KZWH	1	79,472,403	80,454,120	0.98	15
HRFD+KZWH	1	82,197,440	83,194,163	1.00	20
HRFD+KZWH	1	137,951,203	138,624,600	0.67	7
HRFD+KZWH	2	4,241,154	4,314,044	0.07	2
HRFD+KZWH	2	61,033,574	64,011,791	2.98	44
HRFD+KZWH	2	67,254,215	67,762,406	0.51	7
HRFD+KZWH	2	68,396,106	73,711,540	5.32	62
HRFD+KZWH	2	106,230,731	106,382,004	0.15	4
HRFD+KZWH	3	29,706,025	29,706,025	–	2
HRFD+KZWH	3	109,952,877	110,571,030	0.62	12
HRFD+KZWH	4	26,109,668	26,235,250	0.13	3
HRFD+KZWH	4	40,816,726	45,708,458	4.89	58
HRFD+KZWH	4	55,119,203	56,022,265	0.90	5
HRFD+KZWH	4	56,212,197	56,500,795	0.29	10
HRFD+KZWH	4	94,623,036	95,250,718	0.63	11
KALMYK	5	75,347,213 ¹	77,002,200	1.65	29
HRFD+KZWH	5	75,557,340 ¹	75,780,030	0.22	4
RED_STEPPE+MONG	5	75,715,294 ¹	76,942,872	1.23	24
HRFD+KZWH	5	76,111,431	77,042,374	0.93	19
HRFD+KZWH	5	77,410,429	80,040,589	2.63	28
HRFD+KZWH	5	107,756,962	107,861,021	0.10	3
HRFD+KZWH	5	115,381,438	115,960,093	0.58	8
HRFD+KZWH	6	60,054,431	60,471,218	0.42	7

Group	BTA	Beginning (bp)	End (bp)	Length (Mb)	Number of SNPs
HRFD+KZWH	6	65,067,130	81,790,910	16.72	178
KALMYK	7	56,709,192	57,153,518	0.44	12
HRFD+KZWH	8	49,055,911	50,238,146	1.18	22
RED_STEPPE+MONG	8	59,101,606	59,567,379	0.47	11
RED_STEPPE+MONG	8	59,595,738	59,781,489	0.19	1
HRFD+KZWH	8	61,235,958	61,366,316	0.13	3
HRFD+KZWH	8	61,590,116	61,590,116	–	1
HRFD+KZWH	9	43,442,988	44,295,368	0.85	8
HRFD+KZWH	9	62,796,313	62,842,591	0.05	3
HRFD+KZWH	9	62,923,369	62,968,960	0.05	2
HRFD+KZWH	10	14,541,979	14,541,979	–	1
HRFD+KZWH	10	18,170,913	19,291,740	1.12	22
HRFD+KZWH	10	52,818,851	52,818,851	–	1
HRFD+KZWH	11	71,354,308	72,487,911	1.13	11
HRFD+KZWH	11	72,743,792	72,855,583	0.11	2
KALMYK	13	53,566,415	54,357,411	0.79	11
HRFD+KZWH	13	81,502,319	82,172,928	0.67	8
HRFD+KZWH	14	23,430,999	23,430,999	–	1
HRFD+KZWH	14	25,117,380	25,117,380	–	1
HRFD+KZWH	14	50,711,849	50,918,264	0.21	2
HRFD+KZWH	16	33,977,376	34,021,124	0.04	2
HRFD+KZWH	16	34,650,666	35,759,339	1.11	21
HRFD+KZWH	16	42,543,244 ²	46,043,615	3.50	29
RED_STEPPE+MONG	16	42,619,901 ²	43,040,172	0.42	4
HRFD+KZWH	18	35,269,352	36,812,511	1.54	18
HRFD+KZWH	21	2,544,485	3,866,539	1.32	16
HRFD+KZWH	22	21,795,079	23,356,265	1.56	18
HRFD+KZWH	24	35,485,517	35,600,744	0.12	4
HRFD+KZWH	26	41,514,592	41,805,841	0.29	4
HRFD+KZWH	26	42,982,393	43,742,133	0.76	13
HRFD+KZWH	26	47,800,443	50,651,652	2.85	26

BTA = *Bos taurus* autosome, HRFD = Hereford, KALMYK = Kalmyk, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, ROH = Runs of homozygosity, SNP = Single-nucleotide polymorphism. All genomic coordinates are reported relative to the ARS-UCD1.2 assembly (GCF_002263795.3). ¹Overlapping ROH region identified in all three cattle groups. ²Overlapping ROH region identified in the HRFD+KZWH and RED_STEPPE+MONG groups.

The specialized beef group had 51 ROH islands distributed across 19 autosomes. BTA1 contained the largest number, with seven islands, followed by BTA2, BTA4, and BTA5, each containing five. Based on the values presented in Table 9, ROH island lengths in this group ranged from approximately 0.04 to 16.72 Mb, excluding loci for which no measurable interval was represented. Collectively, these regions contained 808 SNPs.

Four ROH islands were identified in the combined steppe group: one on BTA5, two on BTA8, and one on BTA16 (Figure 7). Their lengths ranged from 0.19 to 1.23 Mb, and together they contained 40 SNPs. Three ROH islands were identified in KALMYK, on BTA5, BTA7, and BTA13, containing a total of 52 SNPs. Their reported lengths ranged from 0.44 to 1.65 Mb.

Notably, the BTA5 region was represented in all three groups. The Kalmyk ROH Island extended from 75.35 to 77.00 Mb, the specialized beef group contained overlapping ROH intervals within this region, and the steppe group island extended from 75.72 to 76.94 Mb. In contrast, the BTA7 (56.71–57.15 Mb) and BTA13 (53.57–54.36 Mb) ROH islands were identified only in KALMYK among the groups analyzed.

Functional annotation of genes located within the ROH islands generated 62 annotation clusters across the studied groups, of which eight met the predefined enrichment score threshold of ≥ 1.3 .

Cross-referencing the candidate ROH islands and F_{ST} outlier regions with Cattle QTLdb identified overlaps with previously reported QTLs associated with meat quality and reproductive traits. In particular, the BTA5 region at approximately 75.3–77.0 Mb overlapped the reported marbling QTL (QTL ID 281739) and tenderness QTL (QTL ID 225766). The supplied analysis further identified *USP18* within the BTA5 candidate region and *RNASEH2B* within the shared BTA12 F_{ST} region. These overlaps provide additional biological context for the candidate regions but do not, by themselves, establish causal relationships between the identified loci and the corresponding phenotypes.

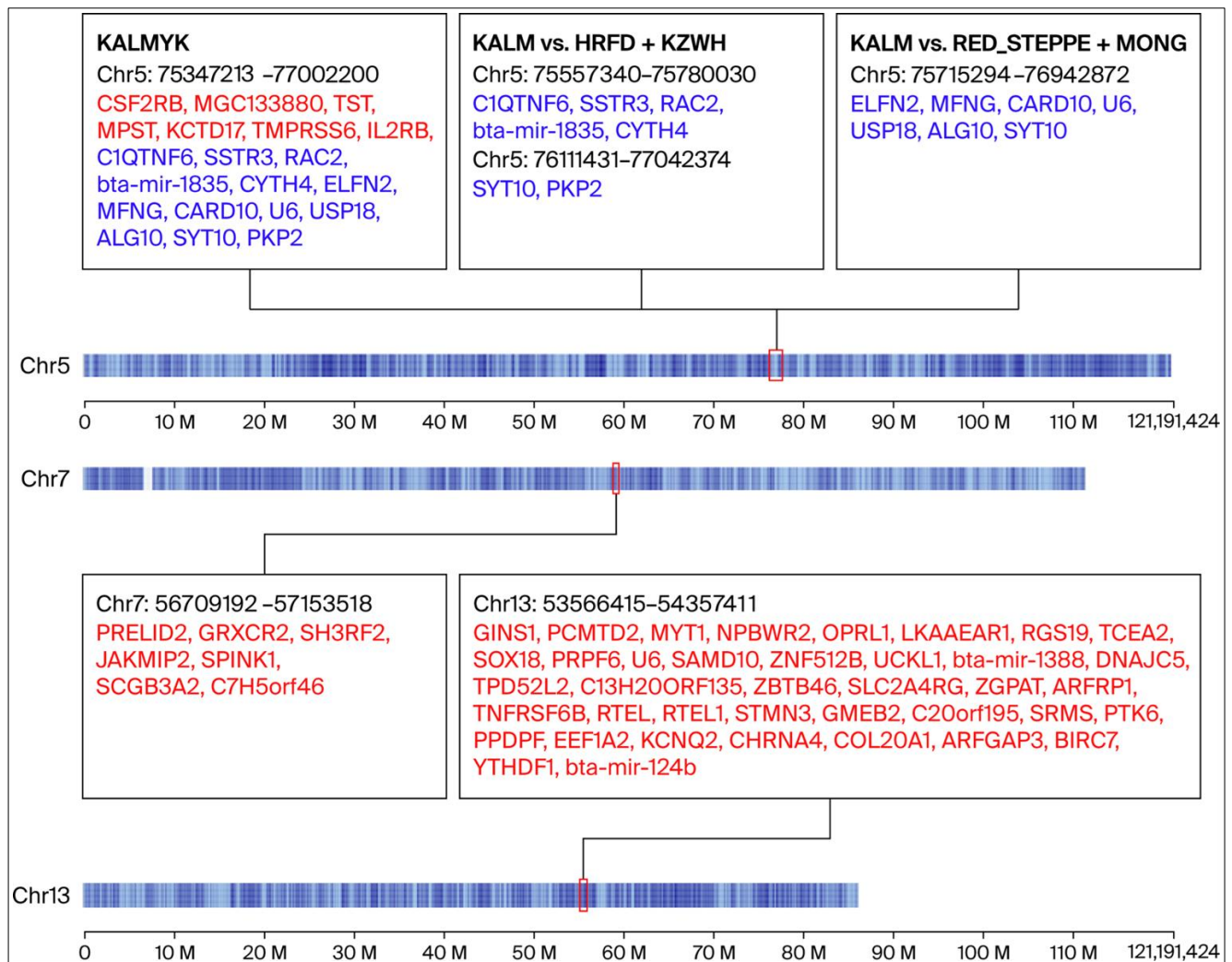


Figure 7: Genes localized within ROH islands identified in Kalmyk cattle. KALMYK = Kalmyk cattle, HRFD+KZWH = Combined specialized beef group comprising Hereford and Kazakh White-Headed cattle, RED_STEPPE+MONG = Combined steppe group comprising Red Steppe and Mongolian cattle, BTA = *Bos taurus* autosome, ROH = Runs of homozygosity. Red indicates genes identified within the corresponding region, whereas blue indicates genes associated with regions shared among multiple groups. All genomic coordinates are reported relative to the ARS-UCD1.2 assembly (GCF_002263795.3).

Analysis of haplotype interpopulation homozygosity (XP-EHH)

XP-EHH analysis was performed to identify genomic regions showing differential haplotype homozygosity between KALMYK and the specialized beef and steppe comparison groups. Candidate SNPs were identified using the empirical threshold applied to the standardized XP-EHH distribution, followed by Benjamini-Hochberg FDR correction. The individual loci retained after Benjamini-Hochberg FDR correction are presented in Table 10. All SNPs listed in Table 10 had FDR-adjusted q-values <0.05. Because a two-sided p-value was used for FDR correction, both extreme positive and negative standardized XP-EHH values could theoretically qualify. However, our biological focus was on selection within the Kalmyk breed; therefore, we primarily interpret positive XP-EHH signals as evidence of selection in the Kalmyk population. The single negative value (BTA12, XP-EHH = -3.74) retained in Table 10 for data completeness and is interpreted as a signal of selection in the reference population.

Table 10: XP-EHH results with FDR-adjusted q-values for candidate loci identified in comparisons of Kalmyk cattle with specialized beef and steppe cattle groups.

Group	BTA	Position (bp)	XP-EHH	p-value	FDR q-value
HRFD+KZWH	1	80,063,107	3.787270	0.000152	0.041210
HRFD+KZWH	1	80,103,951	4.205394	2.61E-05	0.008839
HRFD+KZWH	1	80,896,616	3.870361	0.000109	0.030079
HRFD+KZWH	2	63,112,798	3.922982	8.75E-05	0.026000
HRFD+KZWH	2	63,275,355	4.559049	5.14E-06	0.002335
HRFD+KZWH	2	63,332,645	4.447509	8.69E-06	0.003429

Group	BTA	Position (bp)	XP-EHH	p-value	FDR q-value
HRFD+KZWH	2	63,355,675	4.674024	2.95E-06	0.001513
HRFD+KZWH	2	63,411,488	4.505987	6.61E-06	0.002841
HRFD+KZWH	2	63,460,235	3.941476	8.10E-05	0.024375
HRFD+KZWH	2	63,870,312	3.886185	0.000102	0.028513
HRFD+KZWH	2	63,900,374	4.211825	2.53E-05	0.008714
HRFD+KZWH	2	63,948,276	3.887411	0.000101	0.028513
HRFD+KZWH	4	116,666,797	3.797730	0.000146	0.039958
HRFD+KZWH	5	78,776,781	3.893594	9.88E-05	0.028314
HRFD+KZWH	6	67,643,584	4.195785	2.72E-05	0.009094
HRFD+KZWH	6	67,691,024	3.904369	9.45E-05	0.027408
HRFD+KZWH	6	67,721,544	4.665877	3.07E-06	0.001540
HRFD+KZWH	6	67,924,407	4.412413	1.02E-05	0.003970
HRFD+KZWH	6	67,949,582	4.369424	1.25E-05	0.004762
HRFD+KZWH	6	69,026,906	3.998497	6.37E-05	0.020197
HRFD+KZWH	6	69,184,582	4.661800	3.13E-06	0.001540
HRFD+KZWH	6	69,482,838	4.249907	2.14E-05	0.007573
HRFD+KZWH	6	69,512,310	3.991103	6.58E-05	0.020567
HRFD+KZWH	6	69,639,884	4.493125	7.02E-06	0.002965
HRFD+KZWH	6	69,807,027	4.976273	6.48E-07	0.000473
HRFD+KZWH	6	69,882,331	4.883547	1.04E-06	0.000647
HRFD+KZWH	6	70,000,724	5.041362	4.62E-07	0.000371
HRFD+KZWH	6	70,025,545	6.241011	4.35E-10	1.31E-06
HRFD+KZWH	6	70,065,213	6.671905	2.53E-11	1.52E-07
HRFD+KZWH	6	70,209,132	6.753521	1.44E-11	1.52E-07
HRFD+KZWH	6	70,236,581	6.828341	8.59E-12	1.52E-07
HRFD+KZWH	6	70,320,611	6.115062	9.65E-10	2.11E-06
HRFD+KZWH	6	70,349,791	4.606031	4.10E-06	0.001938
HRFD+KZWH	6	70,376,204	4.939129	7.85E-07	0.000540
HRFD+KZWH	6	70,478,273	4.221860	2.42E-05	0.008456
HRFD+KZWH	6	70,509,911	4.331703	1.48E-05	0.005481
HRFD+KZWH	6	70,927,116	4.748285	2.05E-06	0.001098
HRFD+KZWH	6	70,959,812	3.948766	7.86E-05	0.024251
HRFD+KZWH	6	71,111,080	4.549271	5.38E-06	0.002401
HRFD+KZWH	6	71,154,473	4.473024	7.71E-06	0.003202
HRFD+KZWH	6	71,350,048	5.083342	3.71E-07	0.000319
HRFD+KZWH	6	71,421,017	6.458154	1.06E-10	5.10E-07
HRFD+KZWH	6	71,452,210	6.271917	3.57E-10	1.23E-06
HRFD+KZWH	6	71,519,635	6.696522	2.13E-11	1.52E-07
HRFD+KZWH	6	71,552,977	6.296087	3.05E-10	1.23E-06
HRFD+KZWH	6	71,579,177	5.452179	4.98E-08	7.49E-05
HRFD+KZWH	6	72,051,699	4.320906	1.55E-05	0.005669
HRFD+KZWH	6	72,138,744	4.080028	4.50E-05	0.014653
HRFD+KZWH	6	72,259,859	4.882541	1.05E-06	0.000647
HRFD+KZWH	6	72,301,044	4.883206	1.04E-06	0.000647
HRFD+KZWH	6	72,338,329	5.084432	3.69E-07	0.000319
HRFD+KZWH	6	72,382,208	5.181179	2.20E-07	0.000241
HRFD+KZWH	6	72,454,639	4.977989	6.42E-07	0.000473
HRFD+KZWH	6	72,475,809	4.808316	1.52E-06	0.000873
HRFD+KZWH	6	72,587,575	4.456370	8.34E-06	0.003345
HRFD+KZWH	6	72,837,455	4.812816	1.49E-06	0.000873
HRFD+KZWH	6	72,872,216	4.977301	6.45E-07	0.000473
HRFD+KZWH	6	72,906,612	5.112140	3.19E-07	0.000311
HRFD+KZWH	6	72,939,094	4.350987	1.36E-05	0.005099
HRFD+KZWH	6	73,008,129	4.724475	2.31E-06	0.001208
HRFD+KZWH	6	73,057,109	4.647032	3.37E-06	0.001622
HRFD+KZWH	6	73,092,782	4.523552	6.08E-06	0.002662
HRFD+KZWH	6	73,182,209	5.318808	1.04E-07	0.000132

Group	BTA	Position (bp)	XP-EHH	p-value	FDR q-value
HRFD+KZWH	6	73,212,743	5.334381	9.59E-08	0.000128
HRFD+KZWH	6	73,280,873	5.475680	4.36E-08	7.00E-05
HRFD+KZWH	6	73,307,557	4.766589	1.87E-06	0.001025
HRFD+KZWH	6	73,369,629	4.795336	1.62E-06	0.000909
HRFD+KZWH	6	73,575,770	4.461112	8.15E-06	0.003328
HRFD+KZWH	6	73,639,744	5.102260	3.36E-07	0.000311
HRFD+KZWH	6	73,688,640	5.584893	2.34E-08	4.02E-05
HRFD+KZWH	6	73,819,092	5.106905	3.27E-07	0.000311
HRFD+KZWH	6	73,855,036	5.219673	1.79E-07	0.000206
HRFD+KZWH	6	73,892,735	5.157242	2.51E-07	0.000262
HRFD+KZWH	6	73,932,576	5.360076	8.32E-08	0.000118
HRFD+KZWH	6	74,015,443	4.894241	9.87E-07	0.000647
HRFD+KZWH	6	74,051,400	4.965714	6.84E-07	0.000485
HRFD+KZWH	6	74,140,844	5.656557	1.54E-08	2.86E-05
HRFD+KZWH	6	74,207,957	4.831097	1.36E-06	0.000817
HRFD+KZWH	6	74,246,140	6.003386	1.93E-09	3.88E-06
HRFD+KZWH	6	74,354,100	5.231850	1.68E-07	0.000202
HRFD+KZWH	6	74,378,643	6.141174	8.19E-10	2.09E-06
HRFD+KZWH	6	74,445,066	6.132091	8.67E-10	2.09E-06
HRFD+KZWH	6	74,546,008	5.048327	4.46E-07	0.000370
HRFD+KZWH	6	74,579,431	4.293074	1.76E-05	0.006333
HRFD+KZWH	6	82,409,949	3.945243	7.97E-05	0.024299
HRFD+KZWH	6	82,458,774	4.053782	5.04E-05	0.016181
HRFD+KZWH	6	82,560,990	4.110002	3.96E-05	0.013051
HRFD+KZWH	12	37,025,686	-3.737380	0.000186	0.049752
HRFD+KZWH	20	46,035,562	4.585271	4.53E-06	0.002100
HRFD+KZWH	20	46,066,109	3.918682	8.90E-05	0.026146
RED_STEPPE+MONG	6	106,386,383	4.730721	2.24E-06	0.002374
RED_STEPPE+MONG	7	23,403,749	5.108945	3.24E-07	0.000688
RED_STEPPE+MONG	7	24,625,377	3.939987	8.15E-05	0.043233
RED_STEPPE+MONG	12	77,477,113	4.447041	8.71E-06	0.006719
RED_STEPPE+MONG	16	53,471,014	4.193205	2.75E-05	0.017960
RED_STEPPE+MONG	18	47,995,842	4.783964	1.72E-06	0.002129
RED_STEPPE+MONG	18	51,917,582	4.632172	3.62E-06	0.003236
RED_STEPPE+MONG	18	52,024,379	4.153795	3.27E-05	0.019828
RED_STEPPE+MONG	18	52,486,964	4.621387	3.81E-06	0.003236
RED_STEPPE+MONG	19	36,004,097	5.333242	9.65E-08	0.000273
RED_STEPPE+MONG	19	45,865,387	5.973393	2.32E-09	9.86E-06
RED_STEPPE+MONG	19	45,902,110	6.627853	3.41E-11	2.89E-07
RED_STEPPE+MONG	19	48,209,594	4.985665	6.17E-07	0.001048
RED_STEPPE+MONG	19	52,436,005	4.221402	2.43E-05	0.017175
RED_STEPPE+MONG	19	60,165,029	4.779707	1.76E-06	0.002129
RED_STEPPE+MONG	21	12,712,258	3.987271	6.68E-05	0.037826

HRFD = Hereford, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, BTA = *Bos taurus* autosome, FDR = False discovery rate, XP-EHH = Cross-population extended haplotype homozygosity. The p-values are two-sided values derived from the standardized XP-EHH scores under the normal distribution assumption. FDR q-values are Benjamini–Hochberg-adjusted p-values. All genomic coordinates are reported relative to the ARS-UCD1.2 assembly (GCF_002263795.3).

The XP-EHH analysis revealed a prominent candidate region on BTA6 in the comparison between KALMYK and the combined specialized beef group. This region extended from approximately 67.45 to 73.64 Mb and spanned 6.19 Mb. In the comparison between KALMYK and the combined steppe group, two clustered candidate regions were identified on BTA18 and BTA19, spanning approximately 1.32 and 1.79 Mb, respectively (Figure 8).

Structural annotation of these clustered candidate regions identified 48 genes within the BTA6 region in the comparison between KALMYK and the specialized beef group. In the comparison with the steppe group, 42 genes were identified in the BTA18 region and 36 genes in the BTA19 region (Table 11).

While the XP-EHH analysis identified multiple statistically significant individual SNPs across several chromosomes (Table 10), many of these loci were distributed as isolated signals rather than contiguous clusters. For instance, in the RED_STEPPE+MONG comparison, significant SNPs on BTA6, BTA7, BTA12, BTA16, BTA19

(outside the 44.37-46.16 Mb interval), and BTA21 did not form spatially clustered haplotype blocks meeting the predefined regional criteria. Similarly, in the HRFD+KZWH comparison, significant loci on BTA1, BTA2, BTA4, BTA5, BTA6 (outside the main cluster), BTA12, and BTA20 were isolated. Consequently, these individual SNPs, while statistically significant, were not classified as contiguous candidate regions and were therefore excluded from Table 11, which exclusively reports clustered regional signals.

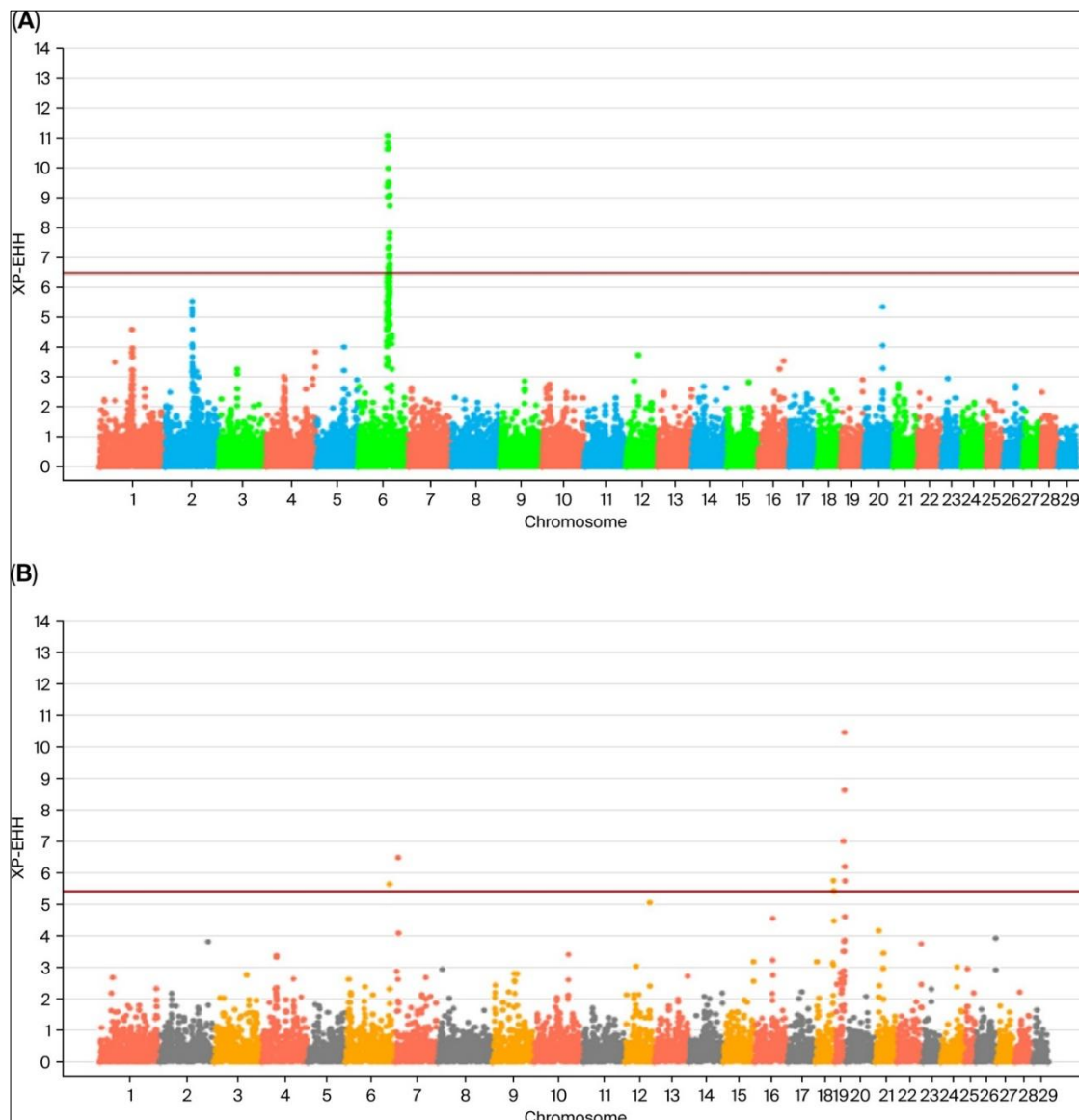


Figure 8: Genome-wide XP-EHH profiles from pairwise comparisons involving Kalmyk cattle. (A) Kalmyk cattle versus the combined specialized beef group comprising Hereford and Kazakh White-Headed cattle. (B) Kalmyk cattle versus the combined steppe group comprising Red Steppe and Mongolian cattle. The x-axis represents *Bos taurus* autosomes, with chromosome widths proportional to their physical lengths, and the y-axis represents XP-EHH values. SNPs are positioned according to their genomic coordinates. The horizontal line represents the empirical threshold applied to identify candidate loci. Colors distinguish adjacent chromosomes and have no additional biological meaning.

Table 11: Candidate genomic regions identified by XP-EHH analysis in comparisons of Kalmyk cattle with specialized beef and steppe cattle groups.

BTA	Group	Beginning (Mb)	End (Mb)	Length (Mb)	Number of SNPs	Genes
6	HRFD+KZWH	67.45	73.64	6.19	69	<i>OCIAD1, OCIAD2, CWH43, DCUN1D4, LRRC66, SGCB, SPATA18, USP46, bta-mir-4449, SNORA26, RASL11B, SCFD2, FIP1L1, LNX1, CHIC2, GSX2, PDGFRA, U6, KIT, KDR, SRD5A3, TMEM165, CLOCK, PDCL2, NMU, EXOC1, CEP135, MGC148692, AASDH, PPAT, PAICS, SRP72, ARL9, HOP, U6atac, SPINK2, MGC157163, NOA1, POLR2B, IGFBP7, EXOC1L, CRACD, THEGL, HOPX, SPINK2B, RESTB, NOA1A, Y_RNA</i>

BTA Group	Beginning (Mb)	End (Mb)	Length (Mb)	Number of SNPs	Genes
18 RED_STEPPE+MONG	51.18	52.50	1.32	4	<i>ZNF574, GRIK5, ATP1A3, RABAC1, ARHGEF1, CD79A, RPS19, DMRTC2, LYPD4, U6atac, CXCL17, CD177, TEX101, BSP3, BSP5, BSP1, LYPD3, PHLDB3, ETHE1, ZNF575, XRCC1, PINLYP, IRGQ, ZNF576, ZNF428, CADM4, PLAUR, IRGC, SMG9, KCNN4, LYPD5, ZNF283, ZNF404, ZNF45, ZNF226, ZNF235, ZNF180, CEACAM20, SNORA70, IGSF23, CEACAM19, CEACAM16</i>
19 RED_STEPPE+MONG	44.37	46.16	1.79	2	<i>C17orf104, CCDC43, DBF4B, ADAM11, GJC1, HIGD1B, EFTUD2, bta-mir-2343, FAM187A, GFAP, KIF18B, C1QL1, DCAKD, NMT1, PLCD3, ACBD4, HEXIM1, HEXIM2, FMNL1, SPATA32, MAP3K14, U6, ARHGAP27, PLEKHM1, RDM1, LYZL6, RPRML, GOSR2, WNT9B, WNT3, NSF, ARF2, CRHR1, SPPL2C, MAPT, KANSL1</i>

HRFD = Hereford, KZWH = Kazakh White-Headed, MONG = Mongolian cattle, RED_STEPPE = Red Steppe, BTA = *Bos taurus* autosome, SNP = Single-nucleotide polymorphism, XP-EHH = Cross-population extended haplotype homozygosity. All genomic coordinates are reported relative to the ARS-UCD1.2 assembly (GCF_002263795.3). Table 11 reports clustered candidate regions only. Individual significant SNPs located outside these clustered intervals (e.g., BTA6 SNPs at ~82.4-88.6 Mb) are presented in Table 10 but were not defined as separate candidate regions because they did not meet the spatial clustering criteria.

The BTA6 region represented the principal clustered XP-EHH signal in the Kalmyk versus specialized beef comparison. This 6.19-Mb interval contained 69 SNPs and 48 annotated genes, including *KIT, KDR, CLOCK, NMU, IGFBP7, PDGFRA, SGCB*, and *SPATA18*. This region also overlapped the broader BTA6 signal identified using the F_ST approach, demonstrating convergence between the two analytical methods within this genomic interval.

In the Kalmyk versus steppe comparison, the clustered BTA18 region extended from 51.18 to 52.50 Mb and contained four SNPs and 42 annotated genes. The BTA19 region extended from 44.37 to 46.16 Mb and contained two SNPs and 36 annotated genes. Although these candidate regions contain genes with diverse biological functions, the present XP-EHH analysis identifies genomic differentiation and does not establish direct genotype–phenotype or causal relationships.

DISCUSSION

Population structure and genetic relationships of KALMYK

Integration of population-genetic and phylogenetic metrics revealed clear relationships among the studied cattle populations. The Neighbor-Net topology supported a close relationship between Hereford and KZWH and, separately, between KALMYK and MONG. These patterns were consistent with the pairwise F_ST estimates, which showed the lowest differentiation for KALMYK/MONG (F_ST = 0.019) and HRFD/KZWH (F_ST = 0.028). The Red Steppe population remained genetically differentiated from these pairs, although shared ancestry components with KALMYK were observed in the ADMIXTURE analysis.

The comparatively low F_ST between KALMYK and MONG is consistent with the reported Turano-Mongolian ancestry of KALMYK [2, 6]. Likewise, the close relationship between Hereford and KZWH is consistent with the documented contribution of HRFD to the development of the KZWH breed. However, the present analyses characterize genomic similarity and differentiation and do not independently establish the historical direction of breed formation.

HRFD showed the lowest A_R among the studied populations. This pattern may be compatible with a history of intensive selection and demographic contraction; however, demographic history was not reconstructed directly in the present study. Therefore, the observed reduction in allelic diversity should be interpreted as a population-genetic characteristic of the sampled Hereford group rather than direct evidence of a specific bottleneck event.

The near-complete distinction between the two sets of F_ST candidate SNPs, with only one shared locus on BTA12, indicates that KALMYK exhibit different patterns of genomic differentiation when compared with specialized beef and steppe cattle. These differences may reflect contrasting demographic histories, founder composition, and selection regimes among the comparison groups, although the present study was not designed to distinguish their relative contributions.

Candidate BTA12 region and *RNASEH2B*

A single F_ST candidate region on BTA12 was shared between both pairwise comparisons. This region contained *RNASEH2B*, and the shared outlier SNP Hapmap30016-BTA-149504 (20,192,311 bp) was located within an intron of this gene.

RNASEH2B encodes a subunit of the RNase H2 complex, which contributes to genome stability by removing ribonucleotides incorporated into DNA and processing RNA–DNA hybrids. In humans, mutations affecting

RNASEH2B are associated with Aicardi–Goutières syndrome and dysregulated interferon responses [42], and the gene has also been implicated in antitumor immunity [43]. Associations with growth-related traits have been reported in chickens [44]. Experimental studies in mice demonstrated that loss of *RNASEH2B* disrupts genome integrity, impairs cell proliferation during embryogenesis, and can result in embryonic lethality [45]. Genome instability associated with RNase H2 dysfunction has also been demonstrated in yeast [46].

In cattle, a deletion on BTA12 spanning the *RNASEH2B* region has been associated with reduced fertility and increased milk yield in European dairy breeds [47]. Comparable deletions affecting fertility have also been described in Norwegian Red and Jersey cattle [48]. In Hanwoo cattle, the *RNASEH2B* region has been associated with total collagen content in the longissimus muscle [49].

Collectively, these findings indicate that the BTA12 region containing *RNASEH2B* is biologically relevant to genomic stability and has previously been associated with productive and reproductive traits in livestock. However, the present study did not evaluate fertility, growth, collagen content, or other phenotypes in the genotyped Kalmyk animals. Therefore, the role of this region in Kalmyk-specific adaptation or reproductive performance remains a hypothesis requiring direct genotype–phenotype validation.

Functional enrichment of differentiating genomic regions

Functional annotation revealed distinct biological themes between the two comparison groups. In the comparison between KALMYK and specialized beef breeds, enriched clusters contained genes related to sexual differentiation, cell adhesion, cellular morphogenesis, cytoskeletal development, and protein-associated processes. In contrast, the steppe comparison yielded an enriched cluster involving signaling pathways related to stem-cell pluripotency, synaptic processes, myopathy-related functions, and circadian regulation.

Circadian regulation pathways are relevant candidates for further investigation because circadian systems influence metabolism, body-temperature regulation, endocrine activity, and responses to photoperiod. These processes may be important in cattle exposed to strong seasonal environmental variation. Nevertheless, the present enrichment analysis identifies functional categories represented among candidate genes and does not demonstrate that these pathways directly mediate the environmental adaptation of KALMYK.

Similarly, enrichment of genes linked to stem-cell pluripotency and tissue-regulatory pathways may indicate differences in developmental or cellular-regulatory processes between the compared populations. Their relationship to resilience under extensive grazing conditions remains speculative and requires functional and phenotypic validation.

ROH patterns and genomic inbreeding

ROH analysis revealed marked differences in genomic inbreeding among the three studied groups. The combined specialized beef group had the highest F_{ROH} (0.133 ± 0.007), whereas KALMYK showed a lower value of 0.050 ± 0.001 and the combined steppe group had an intermediate value of 0.068 ± 0.004 .

The F_{ROH} value observed in KALMYK was broadly comparable with values previously reported for other cattle populations with Turano-Mongolian ancestry, including Yakut cattle [50] and KZWH [51]. The comparatively low F_{ROH} in KALMYK may reflect maintenance of genetic diversity within the breed. However, effective population size, historical mating patterns, and demographic changes were not directly estimated in the present study; therefore, extensive pastoral management should not be considered an established explanation for the observed F_{ROH} value.

The specialized beef group also showed the greatest number and cumulative length of ROH segments, consistent with its higher F_{ROH} . Structural annotation of ROH islands identified 559 genes across 19 autosomes. Most annotated genes were found within ROH islands of the specialized beef group, whereas smaller proportions were identified in the steppe and Kalmyk groups. These differences reflect the distribution and extent of homozygous genomic regions in the studied populations but should not be interpreted as direct quantitative measures of selection pressure without additional demographic and selection analyses.

A BTA5 ROH region was shared across all three groups, whereas an overlapping BTA16 region was detected in the specialized beef and steppe groups. In KALMYK, additional ROH islands were detected on BTA7 and BTA13.

Candidate genes within the shared BTA5 ROH region

The overlapping BTA5 region contained several candidate genes with previously reported functions related to reproduction, metabolism, immunity, and production traits. These included *C1QTNF6*, *SSTR3*, *RAC2*, *bta-mir-1835*, *CYTH4*, *ELFN2*, *MFNG*, *CARD10*, *U6*, *USP18*, *ALG10*, *SYT10*, and *PKP2*.

C1QTNF6 has been proposed as a candidate gene associated with follicular development and reproductive function [52, 53]. Polymorphisms in *C1QTNF6* have also been associated with meat tenderness, marbling, and fat-related traits in cattle [54]. *SSTR3* participates in calcium signaling and regulation of insulin secretion [55] and has been linked to carbohydrate metabolism in buffaloes [56]. *RAC2* encodes a Rho-family GTPase involved in immune-cell signaling, phagocytosis, and oxidative antimicrobial responses [57], and associations with calcium-related processes and parasite resistance have also been reported [58, 59].

CYTH4 has been associated with milk fat content in sheep [60]. *ELFN2* has been implicated in neuronal function and behavioral phenotypes [61], whereas *MFNG* participates in Notch signaling and is involved in cellular proliferation and apoptosis [62]. *CARD10* contributes to inflammatory and immune signaling pathways. *USP18* is involved in innate immunity and interferon-related signaling and has also been implicated in muscle differentiation [63]. *ALG10* has been associated with oocyte maturation [64], whereas *SYT10* has been linked to productive longevity in dairy cattle [65]. Variants in *PKP2* are associated with cardiac function and arrhythmogenic cardiomyopathy [66].

Several of these genes have also been associated with meat quality and reproductive traits in Nellore cattle [67, 68], while candidate associations with milk composition have been reported in buffaloes [69]. These findings provide biological context for the shared BTA5 region. Nevertheless, because the same ROH region was present in all three studied groups, it should not be interpreted as a Kalmyk-specific genomic signature.

XP-EHH candidate regions

XP-EHH analysis identified a prominent BTA6 region in the comparison between KALMYK and the specialized beef group and candidate regions on BTA18 and BTA19 in the comparison with the steppe group.

The BTA6 region contained 48 annotated genes and overlapped a major F_{ST} signal. Genes in this region included *KDR*, *SRD5A3*, *CLOCK*, *NMU*, *IGFBP7*, *KIT*, *PDGFRA*, *SGCB*, and *SPATA18*. Several have previously been implicated in biological processes relevant to vascular development, circadian regulation, metabolism, reproduction, growth, or meat quality traits. However, these literature-derived functions should be viewed as hypotheses for prioritizing candidate genes rather than direct evidence that the same genes influence those traits in KALMYK.

In the comparison with the steppe group, the BTA18 candidate region contained 42 genes and the BTA19 region contained 36 genes. Functional annotation identified clusters related to sperm capacitation, heparin binding, lipoprotein metabolism, fertilization, and regulation of DNA transcription. These pathways suggest potential reproductive and metabolic relevance, but their contribution to breed differentiation remains to be experimentally established.

Convergence among F_{ST}, ROH, and XP-EHH analyses

Comparison of the three analytical approaches identified an important region of convergence on BTA6. The interval from approximately 67.45 to 73.64 Mb was identified using both F_{ST} and XP-EHH in the comparison between KALMYK and the specialized beef group.

Within this region, eight F_{ST} outlier SNPs were identified: ARS-BFGL-NGS-20354 (67,835,794 bp), Hapmap38628-BTA-76690 (67,991,798 bp), Hapmap60836-rs29027147 (68,560,206 bp), Hapmap38134-BTA-107929 (68,700,887 bp), BTA-76626-no-rs (69,310,878 bp), ARS-BFGL-NGS-7423 (70,673,035 bp), ARS-BFGL-NGS-117025 (71,951,614 bp), and BTA-76842-no-rs (73,643,517 bp). The overlapping region contained 25 candidate genes represented in the F_{ST} annotation, including *KDR*, *SRD5A3*, *CLOCK*, *NMU*, and *IGFBP7*.

The convergence of two independent population-comparison approaches supports prioritization of this BTA6 region for further investigation. However, convergence between F_{ST} and XP-EHH does not by itself demonstrate that the region underlies adaptation to continental climate or meat production characteristics. Direct association with phenotypic traits and replication in independent populations are required.

No corresponding overlapping F_{ST} and XP-EHH region was identified in the Kalmyk versus steppe comparison. Moreover, the Kalmyk ROH islands on BTA5, BTA7, and BTA13 did not overlap the principal F_{ST} or XP-EHH candidate regions. This lack of overlap likely reflects the fact that ROH, F_{ST}, and XP-EHH capture different features of genomic variation and selection history, although the precise evolutionary processes responsible cannot be resolved from the present data alone.

Consolidated candidate genes

The principal candidate genes identified through F_{ST}, ROH, and XP-EHH analyses are summarized in Table 12 [38, 47–49, 52–58, 60, 62–96].

Table 12: Consolidated information on prioritized candidate genes identified using F_ST, ROH, and XP-EHH analyses.

Gene symbol	Chromosome	Detection method(s)	Reported biological relevance	Model
<i>RNASEH2B</i>	BTA12	F_ST in both comparisons	RNase H2 complex; genome stability and removal of ribonucleotides from DNA; fertility-related associations reported in cattle [47–49]	Cattle
<i>DCUN1D4</i>	BTA6	F_ST and XP-EHH	Associated with carcass-related traits [38]	Cattle
<i>LRRC66</i>	BTA6	F_ST and XP-EHH	Reported associations with production/reproductive traits in goats and innate immunity [70, 71]	Goat, human
<i>SGCB</i>	BTA6	F_ST and XP-EHH	Muscle development and meat-tenderness-related functions [72]	Cattle
<i>SPATA18</i>	BTA6	F_ST and XP-EHH	Spermatogenesis and semen-quality-related functions [73]	Livestock
<i>bta-mir-4449</i>	BTA6	F_ST and XP-EHH	Reported differential expression in pregnancy-related studies [74, 75]	Cattle, human
<i>SNORA26</i>	BTA6	F_ST and XP-EHH	Putative association with mastitis-related processes [76]	Cattle
<i>RASL11B</i>	BTA6	F_ST and XP-EHH	Cellular signaling and cytoskeletal processes [77]	Pig
<i>SCFD2</i>	BTA6	F_ST and XP-EHH	Cellular trafficking-related functions; reported in cancer-related studies [78]	Cattle, human
<i>CLOCK</i>	BTA6	F_ST and XP-EHH	Circadian regulation and reported association with semen quality [79, 80]	Cattle, boar
<i>PDCL2</i>	BTA6	F_ST and XP-EHH	Spermatogenesis-related functions [81]	Cattle
<i>NMU</i>	BTA6	F_ST and XP-EHH	Reported association with heat tolerance [82]	Cattle
<i>U6</i>	BTA6	F_ST and XP-EHH	Located near loci associated with carcass traits [83]	Cattle
<i>KDR</i>	BTA6	F_ST and XP-EHH	Vascular and melanogenesis-related pathways [84]	Cattle
<i>SRD5A3</i>	BTA6	F_ST and XP-EHH	Steroid and polyprenol-related metabolism [85]	Human
<i>RESTB</i>	BTA6	F_ST and XP-EHH	Transcriptional regulation and neurological processes [86]	Human
<i>NOA1A</i>	BTA6	F_ST and XP-EHH	Mitochondrial and metabolic processes [87]	Cattle
<i>POLR2B</i>	BTA6	F_ST and XP-EHH	Core RNA polymerase II function [88]	Cattle
<i>IGFBP7</i>	BTA6	F_ST and XP-EHH	Reported association with fat deposition and marbling [89]	Cattle
<i>PPAT</i>	BTA6	F_ST and XP-EHH	Purine metabolism and reported association with growth and muscle development [90]	Cattle
<i>PAICS</i>	BTA6	F_ST and XP-EHH	Purine biosynthesis and cellular proliferation [91]	Human
<i>SRP72</i>	BTA6	F_ST and XP-EHH	Signal-recognition-particle function [92]	Human
<i>ARL9</i>	BTA6	F_ST and XP-EHH	Reported association with spermatogenesis [93]	Cattle
<i>HOP</i>	BTA6	F_ST and XP-EHH	Developmental regulation [94]	Human
<i>U6atac</i>	BTA6	F_ST and XP-EHH	Minor spliceosome component; reported association with milk fat yield [95]	Cattle
<i>SPINK2</i>	BTA6	F_ST and XP-EHH	Spermatogenesis-related functions [96]	Mammals
<i>USP18</i>	BTA5	ROH	Innate immunity, interferon regulation, muscle differentiation, and reported association with marbling [63, 67]	Cattle, human
<i>C1QTNF6</i>	BTA5	ROH	Reproductive function and reported associations with marbling, tenderness, and fat deposition [52–54]	Cattle, chicken, pig
<i>SSTR3</i>	BTA5	ROH	Calcium signaling, insulin regulation, and carbohydrate metabolism [55, 56, 69]	Rat, buffalo
<i>RAC2</i>	BTA5	ROH	Immune signaling and reported associations with fat-related traits [57, 58, 69]	Cattle, buffalo
<i>SYT10</i>	BTA5	ROH	Productive longevity, reproduction, and meat quality-related associations [65, 67, 68]	Cattle
<i>ALG10</i>	BTA5	ROH	Reproductive and oocyte-related processes [64, 68]	Cattle, pig
<i>ELFN2</i>	BTA5	ROH	Neuronal function and reported association with meat marbling and lipid metabolism [54, 67]	Cattle
<i>MFNG</i>	BTA5	ROH	Notch signaling, cell proliferation, and apoptosis [62, 67]	Cattle, human
<i>CARD10</i>	BTA5	ROH	Immune/inflammatory signaling and reported production-trait associations [63, 67, 69]	Cattle, human, buffalo
<i>PKP2</i>	BTA5	ROH	Cell–cell adhesion, cardiac function, and reported reproductive associations [66, 68]	Cattle, mouse
<i>CYTH4</i>	BTA5	ROH	Reported association with milk fat content [60, 69]	Buffalo, sheep

BTA = *Bos taurus* autosome, F_ST = Fixation index, ROH = Runs of homozygosity, XP-EHH = Cross-population extended haplotype homozygosity.

Study limitations and future directions

Several limitations should be considered when interpreting the findings. First, the Hereford reference group was relatively small ($n = 27$), which may increase uncertainty in allele frequency estimates and F_ST values. Although the combined specialized beef group comprised 120 animals, no formal statistical power analysis was performed to determine the minimum sample size required for selection-signature detection.

Second, the BovineSNP50 platform was developed largely from variants discovered in European taurine

cattle. Consequently, ascertainment bias may affect characterization of Kalmyk and MONG of Turano-Mongolian ancestry, particularly by underrepresenting rare or population-specific variants. Restricting the analysis to SNPs common across the genotyping platforms and avoiding genotype imputation improved comparability among datasets but does not eliminate this potential bias.

Third, the upper 0.1% empirical F_{ST} threshold was used as an exploratory criterion rather than a formal inferential test. Consequently, the 31 SNPs identified in each comparison should be considered candidate differentiation markers requiring independent validation. Similarly, although XP-EHH signals were supported after FDR correction, the definition of candidate genomic regions depends on the clustering procedure used.

Fourth, no phenotype records for meat quality, fertility, thermotolerance, growth, or other adaptive traits were available for the genotyped animals. Therefore, the biological interpretation of candidate genes relies primarily on functional annotation and previously published associations in other cattle breeds or species. Direct genotype–phenotype associations cannot be inferred from the present study.

Fifth, the use of a medium-density SNP array limits detection of rare variants and fine-scale resolution of candidate genomic regions and ROH boundaries. Higher-density genotyping and whole-genome sequencing would provide more complete variant discovery and improve fine mapping.

Future studies should prioritize independent validation of the 61 candidate SNPs in additional purebred Kalmyk populations and, where appropriate, crossbred populations. Candidate-marker performance should be quantified using measures such as allele frequency differentiation, assignment accuracy, sensitivity, specificity, and cross-validation before application in breed authentication. Integration of genomic data with phenotypic records for growth, meat quality, fertility, reproductive performance, and thermal tolerance would enable direct evaluation of genotype–phenotype relationships. Whole-genome sequencing would further refine the BTA6, BTA12, and other candidate regions and may identify causal or functional variants not captured by the current SNP array.

CONCLUSION

This study provides a comprehensive genome-wide characterization of the population structure, genetic diversity, genomic inbreeding, and selection signatures of KALMYK in comparison with specialized beef and steppe cattle populations. Population-genetic analyses demonstrated close genetic relationships between KALMYK and MONG and between Hereford and KZWH, whereas Red Steppe cattle showed a distinct genetic profile. KALMYK exhibited comparatively low genomic inbreeding ($F_{ROH} = 0.050 \pm 0.001$). Integration of F_{ST} , ROH, and XP-EHH analyses identified several candidate genomic regions associated with population differentiation and potential selection. A shared F_{ST} candidate region on BTA12 contained *RNASEH2B*, with the intronic SNP Hapmap30016-BTA-149504 (20,192,311 bp) differentiating KALMYK from both comparison groups. Furthermore, the BTA6 region at approximately 67.45–73.64 Mb was independently identified by F_{ST} and XP-EHH in the comparison between KALMYK and specialized beef cattle. This region contained candidate genes including *CLOCK*, *KDR*, *NMU*, *SRD5A3*, and *IGFBP7*. ROH analysis also identified a BTA5 region shared among the studied groups containing genes related to reproductive, metabolic, immune, and production-associated processes.

The genomic regions and candidate markers identified in this study provide a foundation for genetic characterization and monitoring of KALMYK. In particular, the shared BTA12 differentiation signal and the convergent F_{ST} and XP-EHH signal on BTA6 represent priority regions for further investigation. Following independent validation, highly informative SNPs from these regions and other identified loci could contribute to a reduced marker panel for breed authentication and genetic monitoring. Such markers may also support conservation programs by facilitating assessment of breed genetic integrity and informing breeding decisions aimed at maintaining the distinctive genetic resources of KALMYK.

A major strength of this study is the integration of complementary population-genomic approaches, including population structure analysis, F_{ST} , ROH, F_{ROH} , and XP-EHH, which enabled the genomic characteristics of KALMYK to be examined from multiple perspectives. The study included 100 purebred KALMYK from the breed's core breeding region, substantially increasing the genomic representation of this population compared with previously reported studies. The use of an identical set of 44,774 SNPs across the comparison groups reduced inconsistencies arising from differences among SNP chip versions. Avoiding genotype imputation further ensured that comparisons were based on directly observed genotypes. Importantly, convergence between F_{ST} and XP-EHH independently prioritized the BTA6 region, strengthening its relevance as a candidate region for subsequent validation.

Several limitations should be considered when interpreting these findings. The Hereford sample was relatively small ($n = 27$), which may increase uncertainty in allele frequency and F_{ST} estimates, although the combined HRFD+KZWH comparison group included 120 animals. No formal statistical power analysis was conducted to determine the minimum sample size required for selection-signature detection. In addition, the BovineSNP50 platform was developed primarily from variants identified in European taurine cattle and may therefore underrepresent rare or population-specific variants in cattle of Turano-Mongolian ancestry. The top 0.1% empirical F_{ST} threshold represents an exploratory criterion rather than a formal inferential test; consequently, the identified F_{ST} outliers should be regarded as candidate differentiation markers. Furthermore, phenotypic records for meat quality, growth, fertility, reproductive performance, and thermotolerance were unavailable for the genotyped animals, preventing direct genotype–phenotype association analysis. Finally, the medium-density SNP array provides limited resolution for fine mapping and may not capture rare variants or precisely define the boundaries of candidate regions and ROH islands.

Future studies should validate the 61 candidate SNPs in independent KALMYK populations using targeted genotyping approaches and assess their discriminatory performance for breed authentication. Subsequent analyses should determine whether the candidate set can be reduced to a core panel of approximately 20–30 highly informative markers without substantially reducing classification accuracy. Integration of genomic information with phenotypic records for growth, meat quality, fertility, reproductive performance, and heat and cold tolerance will be essential to establish direct genotype–phenotype relationships. Higher-density genotyping and whole-genome sequencing should also be used to refine the candidate regions, particularly the prioritized BTA6 and BTA12 loci, identify additional population-specific variants, and improve resolution of ROH boundaries. Functional investigation of prioritized genes, including *RNASEH2B*, *CLOCK*, *KDR*, *NMU*, *IGFBP7*, *USP18*, and *C1QTNF6*, may further clarify their potential contribution to the biological characteristics of KALMYK.

The present study expands the genomic characterization of purebred KALMYK and identifies distinct patterns of population structure, genomic inbreeding, and genomic differentiation relative to specialized beef and steppe cattle. The BTA12 region containing *RNASEH2B* and the convergent F_{ST} –XP-EHH signal on BTA6 represent particularly important candidates for further investigation. Together with the ROH regions and additional candidate loci identified in this study, these findings establish a genomic framework for future functional validation, breed authentication, genetic monitoring, and conservation of KALMYK. Independent validation and integration with phenotypic and whole-genome sequence data will be essential before these candidate markers can be incorporated into routine breeding or conservation programs.

DATA AVAILABILITY

The raw genotype data supporting the findings of this study, comprising Illumina BovineSNP50 v3 genotyping data from 100 Kalmyk cattle, are available through the Zenodo repository (Doi: 10.5281/zenodo.22159534).

GENERATIVE AI DECLARATION

AI-generated images were used solely for illustrative purposes in the graphical abstract. Their use did not affect the scientific data, analyses, interpretation, or conclusions of this study.

AUTHORS' CONTRIBUTIONS

ZB and NZ: Conceptualization and writing – review and editing. BS and KA: Data curation. AA and TD: Formal analysis, methodology, and software. BB and NC: Formal analysis and methodology. AU: Sample collection. AA, ZB, and TD: Validation and investigation. AA, VU, and ZB: Writing – original draft. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

This study was conducted within the framework of the State Assignment of the Ministry of Science and Higher Education of the Russian Federation (No. 075-03-2026-375/4), entitled “Search for genetic markers specific to Kalmyk cattle, with the aim of subsequently developing a panel of reference polymorphisms (‘genetic passport of the Kalmyk breed’).”

COMPETING INTERESTS

The authors declare that they have no competing interests.

PUBLISHER'S NOTE

Veterinary World remains neutral with regard to jurisdictional claims in published institutional affiliations.

REFERENCES

1. Zinovieva NA, Deniskova TE, Kharzinova VR, Bagirov VA, Romanov MN, Volkova VV, *et al.* Conservation of native livestock breeds in Russia: current state and promising prospects. *Animals*. 2025;15:3103.
2. Abdelmanova AS, Kharzinova VR, Volkova VV, Dotsev AV, Sermyagin AA, Boronetskaya OI, *et al.* Comparative study of the genetic diversity of local steppe cattle breeds from Russia, Kazakhstan and Kyrgyzstan by microsatellite analysis of museum and modern samples. *Diversity*. 2021;13:351.
3. Sleptsov II, Tarabukhin NI, Miroshnikov SA, Frolov AN, Zavyalov OA, Balakirev NA. Adaptive changes of the elemental status of Kalmyk cattle to conditions of biogeochemical province of the Republic of Sakha (Yakutia). *IOP Conf Ser Earth Environ Sci*. 2019;341:012039.
4. Bekbolotova AT, Aitzhanova IN, Naimanov DK, Shamshidyn AS, Dzhulamanov KM. The influence of the slaughter age on the meat productivity of Kalmyk cattle. *GBJ*. 2023;3(72):35–42.
5. Bozymov K, Kosilov V, Shakirbek N. Characteristics of growth and development of young Kalmyk cattle. *GBJ*. 2023;1(2)(71):92–100.
6. Sermyagin AA, Dotsev AV, Gladyr EA, Traspov AA, Deniskova TE, Kostyunina OV, *et al.* Whole-genome SNP analysis elucidates the genetic structure of Russian cattle and its relations with Eurasian taurine breeds. *Genet Sel Evol*. 2018;50:37.
7. Aitzhanova IN, Shaikamal GI, Seleuova LA, Gabdullin SS, Bekbolotova AT. Comparative assessment of meat qualities of purebred and crossbred Kalmyk bulls. *OnLine J Biol Sci*. 2022;22(3):395–403.
8. Pristupa VN, Svyatogorov NA, Torosyan DS, Svitenko OV, Svyatogorova AE. Indicators of meat productivity of dairy and meat breeds. *Proc Kuban State Agr Univ*. 2023;(109):171–177.
9. Kolosov YA, Pristupa VN, Koshchayev AG, Eremenko ON, Nesterenko AA. Meat productivity of bull calves under intensive rearing. *Sci J KubSAU*. 2020;(159):183–194.
10. Zinovieva NA, Dotsev AV, Sermyagin AA, Deniskova TE, Abdelmanova AS, Kharzinova VR, *et al.* Selection signatures in two oldest Russian native cattle breeds revealed using high-density single nucleotide polymorphism analysis. *PLoS One*. 2020;15:e0242200.
11. Abdelmanova AS, Sermyagin AA, Dotsev AV, Rodionov AN, Stolpovsky YA, Zinovieva NA. Whole-genomic studies of the population structure of Russian local Black-Pied breeds. *Russ J Genet*. 2022;58:804–813.
12. Smith JL, Wilson ML, Nilson SM, Rowan TN, Schnabel RD, Decker JE, *et al.* Genome-wide association and genotype by environment interactions for growth traits in US Red Angus cattle. *BMC Genomics*. 2022;23:517.
13. Ma J, Gao X, Li J, Gao H, Wang Z, Zhang L, *et al.* Assessing the genetic background and selection signatures of Huaxi cattle using high-density SNP array. *Animals*. 2021;11:3469.
14. Huang N, Zhao L, Wang J, Jiang Q, Ju Z, Wang X, *et al.* Signatures of selection in indigenous Chinese cattle genomes reveal adaptive genes and genetic variations to cold climate. *J Anim Sci*. 2023;101:skad006.
15. Goyache F, Pérez-Pardal L, Fernández I, Traoré A, Menéndez-Arias NA, Álvarez I. Ancient autozygous segments subject to positive selection suggest adaptive immune responses in West African cattle. *Gene*. 2021;803:145899.
16. Jasielczuk I, Gurgul A, Szmatoła T, Radko A, Majewska A, Sosin E, *et al.* The use of SNP markers for cattle breed identification. *J Appl Genet*. 2024;65:575–589.
17. Wang H, Wu H, Zhang W, Jiang J, Qian H, Man C, *et al.* Development and validation of a 5K low-density SNP chip for Hainan cattle. *BMC Genomics*. 2024;25:873.
18. Neumann GB, Korkuć P, Arends D, Wolf MJ, May K, Reißmann M, *et al.* Design and performance of a bovine 200 k SNP chip developed for endangered German Black Pied cattle (DSN). *BMC Genomics*. 2021;22:905.
19. Michiels B, Pardo AM, Masagué MFO, Giovambattista G, Corva PM. Characterization of experimental cattle populations from Argentina with a low-density SNP genotyping panel. *Anim Genet*. 2023;54:68–72.
20. Gorlov IF, Randelin AV, Slozhenkina MI, Natyrov AK, Bolaev BK, Knyaxhenchenko OA, *et al.* Identification of SNP-profiles and evaluation of breeding value of Kalmyk breed cattle using biochips. *Zootech*. 2019;4:9–11.
21. Khamzina AK, Yurchenko AA, Yudin NS, Ibragimov PS, Ussenbekov YS, Larkin DM. History, status and genetic characteristics of native cattle breeds from the Republic of Kazakhstan. *Vavilov J Genet Breed*. 2024;28:416–423.

22. Khamzina AK, Igoshin AV, Muslimova ZU, Turgumbekov AA, Khussainov DM, Yudin NS, *et al.* Resequencing composite Kazakh White-Headed cattle: insights into ancestral breed contributions, selection signatures, and candidate genetic variants. *Animals*. 2025;15:385.
23. Chimidova NV, Moiseykina LG, Ubushieva AV, *et al.* Current state and review of genetic studies of the domestic Kalmyk cattle breed. *Agric Sci J*. 2025;(11):123–129.
24. Chimidova N, Ubushieva A, Ubushieva V, Bochkaeva Z. Genetic polymorphisms influencing meat productivity and quality in Kalmyk cattle of Russia: a systematic review of growth hormone, thyroglobulin, leptin, and calpain-1 genes. *Vet World*. 2026;19(2):523–538.
25. Sempéré G, Moazami-Goudarzi K, Eggen A, Laloë D, Gautier M, Flori L. WIDDE: a web-interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics*. 2015;16:940.
26. Gao Y, Gautier M, Ding X, Zhang H, Wang Y, Wang X, *et al.* Species composition and environmental adaptation of indigenous Chinese cattle. *Sci Rep*. 2017;7:16196.
27. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, *et al.* Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol*. 2003;68:69–78.
28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575.
29. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. *diversity*: an R package for the assessment and exploration of population genetics parameters and their associated errors. *Methods Ecol Evol*. 2013;4:782–788.
30. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978;89:583–590.
31. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38:1358–1370.
32. Kalinowski ST. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conserv Genet*. 2004;5:539–543.
33. Wickham H. *Elegant graphics for data analysis*. Cham: Springer International Publishing; 2016.
34. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006;23:254–267.
35. Alexander DH, Novembre J, Lange K. Fast model-based assessment of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–1664.
36. Milanesi M, Capomaccio S, Vajana E, Bomba L, Garcia JF, Ajmone-Marsan P, *et al.* BITE: an R package for biodiversity analyses. *Bioinformatics*. 2017;181610.
37. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat Rev Genet*. 2009;10:639–650.
38. Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet Sel Evol*. 2015;47:49.
39. Biscarini F, Cozzi P, Gaspa G, Marras G. *detectRUNS*: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. 2018. Available from: <https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html>. Accessed 2026 Apr 25.
40. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007;449:913–918.
41. Gautier M, Klassmann A, Vitalis R. REHH 2.0: an implementation of the R package REHH to detect positive selection from haplotype structure. *Mol Ecol Resour*. 2017;17:78–90.
42. Cristini A, Tellier M, Constantinescu F, Accalai C, Albulescu LO, Heiringhoff R, *et al.* RNase H2, mutated in Aicardi-Goutières syndrome, resolves co-transcriptional R-loops to prevent DNA breaks and inflammation. *Nat Commun*. 2022;13:2961.
43. Maxwell MB, Hom-Tedla MS, Yi J, Li S, Rivera SA, Yu J, *et al.* *ARID1A* suppresses R-loop-mediated STING-type I interferon pathway activation of anti-tumor immunity. *Cell*. 2024;187:3390–3408.e19.
44. Ou JH, Rönneburg T, Carlborg Ö, Ferst Honaker C, Siegel PB, Rubin CJ. Complex genetic architecture of the chicken Growth1 QTL region. *PLoS One*. 2024;19:e0295109.
45. Reijns MA, Rabe B, Rigby RE, Mill P, Astell KR, Lettice LA, *et al.* Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*. 2012;149:1008–1022.

46. McElhinny SAN, Kumar D, Clark AB, Watt DL, Watts B, Lundström EB, *et al.* Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol.* 2010;6:774–781.
47. Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbrandtsen B, Karim L, *et al.* A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red Cattle: additional evidence for the common occurrence of balancing selection in livestock. *PLoS Genet.* 2014;10:e1004049.
48. Wu X, Uddin M, Guldbrandtsen B, Lund MS, Sahana G. Novel haplotypes responsible for prenatal death in Nordic Red and Danish Jersey cattle. *J Dairy Sci.* 2020;103:4570–4578.
49. Won KH, Kim DH, Hwang IH, Lee HK, Oh JD. Genome-wide association studies on collagen contents trait for meat quality in Hanwoo. *J Anim Sci Technol.* 2022;65:311.
50. Mahar K, Goli RC, Chishi KG, Ganguly I, Dixit SP, Singh S, *et al.* Runs of homozygosity decipher genetic diversity in cattle breed dwelling in the colder regions of the world. *Cytogenet Genome Res.* 2024;164(3–4):154–164.
51. Beishova I, Dossybayev K, Shamshidin A, Belaya A, Bissembayev A, Khamzin K, *et al.* Distribution of homozygosity regions in the genome of Kazakh cattle breeds. *Diversity.* 2022;14(4):279.
52. Li C, Li Q, Wu X, Zhang Z, Li J, Jiang R, *et al.* Localization and expression of *C1QTNF6* in chicken follicles and its regulatory effect on follicular granulosa cells. *Poult Sci.* 2025;104:104538.
53. Yin L, Wang W, Wei H, Xi F, Chu G, Yang G. Localization and expression of *CTRP6* in ovary and its regulation by FSH in porcine granulosa cells. *Theriogenology.* 2019;127:56–65.
54. Kim KS, Kim SW, Raney NE, Ernst CW. Evaluation of BTA1 and BTA5 QTL regions for growth and carcass traits in American and Korean cattle. *Asian-Australas J Anim Sci.* 2012;25:1521–1528.
55. Sivcev S, Constantin S, Smiljanic K, Sokanovic SJ, Fletcher PA, Sherman AS, *et al.* Distribution and calcium signaling function of somatostatin receptor subtypes in rat pituitary. *Cell Calcium.* 2024;124:102967.
56. Huang J, Feng X, Zhu R, Guo D, Wei Y, Cao X, *et al.* Comparative transcriptome analysis reports that *PCK1* is a potential gene affecting IMF position in buffalo. *BMC Genomics.* 2020;21:710.
57. Cheng Z, Palma-Vera S, Buggiotti L, Salavati M, Becker F, Werling D, *et al.* Transcriptomic analysis of circulating leukocytes obtained during the recovery from clinical mastitis caused by *Escherichia coli* in Holstein dairy cows. *Animals.* 2022;12:2146.
58. Vieira-Neto A, Poindexter M, Marinho MN, Zimpel R, Husnain A, Silva A, *et al.* Effect of source and amount of vitamin D on function and mRNA expression in immune cells in dairy cows. *J Dairy Sci.* 2021;104:10796–10811.
59. Berton MP, Silva RMdO, Peripolli E, Stafuzza NB, Martin JF, Álvarez MS, *et al.* Genomic regions and paths associated with gastrointestinal parasites resistance in Santa Inês breed adapted to tropical climate. *J Anim Sci Biotechnol.* 2017;8:73.
60. Rezvannejad E, Asadollahpour Nanaei H, Esmailzadeh A. Detection of candidate genes affecting milk production traits in sheep using whole-genome sequencing analysis. *Vet Med Sci.* 2022;8:1197–1204.
61. Dunn HA, Zucca S, Dao M, Orlandi C, Martemyanov KA. *ELFN2* is a postsynaptic cell adhesion molecule with essential roles in controlling group III mGluRs in the brain and neuropsychiatric behavior. *Mol Psychiatry.* 2019;24:1902–1919.
62. Gao Y, Luo L, Qu Y, Zhou Q. *MFNG* is an independent prognostic marker for osteosarcoma. *Eur J Med Res.* 2023;28:256.
63. Olie CS, Pinto-Fernández A, Damianou A, Vendrell I, Mei H, Hamer BD, *et al.* *USP18* is an essential regulator of muscle cell differentiation and mutation. *Cell Death Dis.* 2023;14:231.
64. Wu Z, Mou Q, Fang T, Wang Y, Liang H, Wang C, *et al.* Global 3'-untranslated region landscape mediated by alternative polyadenylation during meiotic maturation of pig oocytes. *Reprod Domest Anim.* 2022;57:33–44.
65. Buaban S, Lengnudum K, Boonkum W, Phakdeedindan P. Genome-wide association study on milk production and somatic cell score for Thai dairy cattle using weighted single-step approach with random regression test-day model. *J Dairy Sci.* 2022;105:468–494.
66. Vencato S, Romanato C, Rampazzo A, Calore M. Animal models and molecular pathogenesis of arrhythmogenic cardiomyopathy associated with pathogenic variants in intercalated disc genes. *Int J Mol Sci.* 2024;25:6208.
67. Magalhães AFB, De Camargo GMF, Fernandes GA, Gordo DGM, Tonussi RL, Costa R, *et al.* Genome-wide association study of meat quality traits in Nellore cattle. *PLoS One.* 2016;11:e0157845.
68. Barreto Amaral Teixeira D, Alves Fernandes GJ, Beraldo dos Santos Silva D, Bernal Costa R, Takada L, Mansan Gordo Tiago DG, *et al.* Genomic analysis of stability in Nellore cattle. *PLoS One.* 2017;12:e0179076.

69. de Camargo GM, Aspilcueta-Borquis RR, Fortes MR, Porto-Neto R, Cardoso DF, Santos DJ, *et al.* Prospecting major genes in dairy buffaloes. *BMC Genomics*. 2015;16:872.
70. Xiong J, Bao J, Hu W, Shang M, Zhang L. Whole-genome resequencing reveals genetic diversity and selection characteristics of dairy goat. *Front Genet*. 2023;13:1044017.
71. Ng AC, Eisenberg JM, Heath RJ, Huett A, Robinson CM, Nau GJ, *et al.* Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proc Natl Acad Sci USA*. 2011;108 Suppl 1:4631–4638.
72. Zhao C, Tian F, Yu Y, Luo J, Mitra A, Zhan F, *et al.* Functional genomic analysis of variation on beef tenderness induced by acute stress in Angus cattle. *Comp Funct Genomics*. 2012;2012:756284.
73. Khan MZ, Chen W, Naz S, Liu X, Liang H, Chen Y, *et al.* Determinant genetic markers of semen quality in livestock. *Front Endocrinol*. 2024;15:1456305.
74. Ross PJ, Mok RS, Smith BS, *et al.* Modeling neuronal consequences of autism-associated gene regulatory variants with human induced pluripotent stem cells. *Mol Autism*. 2020;11:33.
75. Lim HJ, Kim HJ, Lee JH, Lim DH, Son JK, Kim ET, *et al.* Identification of plasma miRNA biomarkers for pregnancy detection in dairy cattle. *J Anim Reprod Biotechnol*. 2021;36:35–44.
76. Omonijo FA, Wang M, Gagné D, Laterrière M, Genier S, Zhao X, *et al.* Small nucleolar RNA dysregulation and potential roles in bovine subclinical mastitis. *J Anim Sci Biotechnol*. 2025;16(1):124.
77. Belous AA, Sermiyagin AA, Zinovieva NA. Genetic assessment of predicted residual feed consumption and expression of significant candidate genes in Duroc pigs and second-generation commercial crossbreeds. *Genetika*. 2023;59:11.
78. Rosewick N, Durkin K, Artesi M, *et al.* Cis-perturbation of cancer drivers by the HTLV-1/BLV proviruses is an early determinant of leukemogenesis. *Nat Commun*. 2017;8:15264.
79. Kumar S, Deb R, Singh U, Ganguly I, Mandal D, Tyagi S, *et al.* Bovine circadian locomotor output cycles kaput (*CLOCK*) and clusterin (*CLU*) mRNA quantitation in ejaculated crossbred bull spermatozoa. *Reprod Dom Anim*. 2015;50:505–509.
80. Earnhardt-San AL, Baker EC, Riley DG, Ghaffari N, Long CR, Cardoso RC, *et al.* Differential expression of circadian clock genes in the bovine neuroendocrine adrenal system. *Genes (Basel)*. 2023;14(11):2082.
81. Zhao Q, Liu R, Miao X, Mei Y, Wang X, Ma J, *et al.* Identification of candidate proteins influencing spermatogenesis in Shandong black cattle via integrated multiomics analysis. *BMC Genomics*. 2025;26(1):628.
82. Katiyatiya CLF, Bradley G, Muchenje V. Thermotolerance, health profile and cellular expression of *HSP90AB1* in Nguni and Boran cows raised on natural pastures under tropical conditions. *J Therm Biol*. 2017;69:85–94.
83. Rafter P, Gormley IC, Purfield D, Parnell AC, Naderi S, Berry DP. Genome-wide association analyses of carcass traits using copy number variants and raw intensity values of single nucleotide polymorphisms in cattle. *BMC Genomics*. 2021;22(1):757.
84. Whitacre L, Decker J, Kim JW, Schnabel R, Taylor J. Kinase insert domain receptor—a candidate for the Hereford “spotted” gene. *Plant Anim Genome XXI*. 2013:0543.
85. Stiles AR, Russell DW. *SRD5A3*: a surprising role in glycosylation. *Cell*. 2010;142(2):196–198.
86. Hwang JY, Zukin RS. REST, a master transcriptional regulator in neurodegenerative disease. *Curr Opin Neurobiol*. 2018;48:193–200.
87. Besnard F, Guintard A, Grohs C, Guzylack-Piriou L, Cano M, Escoufflaire C, *et al.* Massive detection of cryptic recessive genetic defects in dairy cattle mining millions of life histories. *Genome Biol*. 2024;25(1):248.
88. Falchi L, Cesarani A, Brito LF, Mastrangelo S, Paucullo A, Macciotta NPP, *et al.* Runs of homozygosity in Italian Holstein bulls: a permutation approach and time-based mapping of the genomic regions potentially under selection. *BMC Genomics*. 2026;27(1):203.
89. Dai S, Zhao P, Li W, Peng L, Jiang E, Du Y, *et al.* Global pangenome analysis highlights the critical role of structural variants in cattle improvement and identifies a unique event as a novel enhancer in *IGFBP7*+ cells. *Mol Biol Evol*. 2025;42(9):msaf205.
90. Liang M, An B, Deng T, Du L, Li K, Cao S, *et al.* Incorporating genome-wide and transcriptome-wide association studies to identify genetic elements of longissimus dorsi muscle in Huaxi cattle. *Front Genet*. 2023;13:982433.
91. Huo A, Xiong X. *PAICS* as a potential target for cancer therapy linking purine biosynthesis to cancer progression. *Life Sci*. 2023;331:122070.

92. Kirwan M, Walne AJ, Plagnol V, Velangi M, Ho A, Hossain U, *et al.* Exome sequencing identifies autosomal-dominant *SRP72* mutations associated with familial aplasia and myelodysplasia. *Am J Hum Genet.* 2012;90(5):888–892.
93. Gao Y, Bai F, Zhang Q, An X, Wang Z. Dynamic transcriptome profiles and novel markers in bovine spermatogenesis revealed by single-cell sequencing. *J Integr Agric.* 2023.
94. Caspa Gokulan R, Yap LF, Paterson IC. *HOPX*: a unique homeodomain protein in development and tumor suppression. *Cancers (Basel).* 2022;14(11):2764.
95. Zielke LG, Bortfeldt RH, Reissmann M, Tetens J, Thaller G, *et al.* Impact of variation at the *FTO* locus on milk fat yield in Holstein dairy cattle. *PLoS One.* 2013;8(5):e63406.
96. Yang T, Liu H, Yang Y, Hu P, Song X, Peng Y, *et al.* Identification of key genes associated with spermatogenesis arrest in fox hybrids using weighted gene co-expression network analysis. *Theriogenology.* 2020;147.
