Niyazbekova et al. *Zool. Res.* 2025, 46(2): 301−311
https://doi.org/10.24272/j.issn.2095-8137.2024.235
https://cstr.cn/32248.14.zr.2024.235

ZR | Zoological Research

Article Open Access

Whole-genome sequencing reveals genetic architecture and selection signatures of Kazakh cattle

Zhannur Niyazbekova[1,2,6,#], Yuan Xu[1,#], Min Qiu[1], Hao-Ping Wang[1], Ibragimov Primkul[2], Hojjat Asadollahpour Nanaei[1,3], Yessengali Ussenbekov[2], Kuanysh Kassen[4], Yi Liu[5], Cai-Yue Gao[1], Shynar Akhmetsadykova[6], Nuriddin Ruzikulov[7], Yu Jiang[1,*], Yu-Dong Cai[1,*]

[1] *College of Animal Science and Technology*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*
[2] *Faculty of Veterinary Medicine*, *Kazakh National Agrarian Research University*, *Almaty* 050000, *Kazakhstan*
[3] *College of Life Sciences*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*
[4] *College of Plant Protection*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*
[5] *Institute of Animal Husbandry and Veterinary Science*, *Tianjin Academy of Agricultural Sciences*, *Tianjin* 300381, *China*
[6] *Research and Production Enterprise "ANTIGEN" Co. Ltd.*, *Abai Village*, *Almaty* 050409, *Kazakhstan*
[7] *Samarkand State University of Veterinary Medicine*, *Animal Husbandry and Biotechnology*, *Samarkand* 140103, *Uzbekistan*

ABSTRACT

Local cattle breeds play a critical role in breeding programs due to their genetic adaptations to diverse environmental conditions. However, the genomic architecture of local cattle breeds in Kazakhstan remains largely unexplored. This study utilized whole-genome sequencing data from Kazakh cattle to elucidate their genetic composition, uncovering three primary ancestral components: European, Eurasian, and East Asian taurine. The East Asian taurine lineage likely represents the earliest genetic contribution to Kazakh cattle but was largely replaced by subsequent waves of cattle migrations across Eurasia, leaving only a minor genetic signature in the current cattle population. In contrast, Eurasian taurine ancestry predominated in the Alatau and Kazakh local breeds, while the European taurine component was most prevalent in Kazakh white-headed cattle, consistent with their documented breeding history. Kazakh cattle exhibited higher genetic diversity and lower inbreeding coefficients compared to European commercial breeds, reflecting reduced exposure to intense artificial selection. A strong selection signal was identified on chromosome 6 at a locus encompassing *PDGFRA*, *KIT*, and *KDR*, which may be associated with the white-headed pigmentation characteristic of Kazakh white-headed cattle. Additional genes under selection were linked to lipid metabolism (*IRS1*, *PRKG1*, and *ADCY8*), meat production traits (*KCNMA1*, *PDGFRA*, *HIF1A*, and *ANTXR1*), and dairy production (*ATP2B1*, *DHX15*, *FUK*, *NEGR1*, *CCDC91*, *COG4*, and *PTK2B*). This study represents the first comprehensive analysis of nuclear genome data from local Kazakh cattle. It highlights the impact of historical cattle migrations across Eurasia on their genetic landscape and identifies key genomic regions under selection. These findings advance our understanding of the evolutionary history of cattle and offer valuable genetic resources for future breeding strategies.

Keywords: Kazakh white-headed cattle; Alatau cattle; Local breed; Genetic diversity; Whole-genome sequencing; Selection signatures.

INTRODUCTION

The domestication of cattle (*Bos taurus*) began approximately 10 000 years ago, with two primary centers identified (Loftus et al., 1994; Vigne et al., 2005), notably the Near East for *Bos taurus* and the Indus Valley for *Bos indicus* (Loftus et al., 1994). Over time, cattle domestication has led to the development of a thousand breeds worldwide, each adapted to diverse environmental conditions. Many of these breeds have been commercially developed through extensive crossbreeding efforts (Leroy et al., 2016). Global genome-wide studies categorize domestic cattle into five major continental groups: European taurine, Eurasian taurine, East Asian taurine, Chinese indicine, and Indian indicine (Chen et al., 2018a). Local cattle breeds are increasingly important in developing countries, where their adaptability to local

Received: 01 October 2024; Accepted: 03 December 2024; Online: 04 December 2024

Foundation items: This work was supported by the National Key R&D Program of China (2022YFF1000100) and China Postdoctoral Science Foundation (2022M722615)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: yu.jiang@nwafu.edu.cn; yudong.cai@nwafu.edu.cn

environments makes them valuable sources of high-quality agricultural products. These breeds often exhibit strong resistance to systemic infectious diseases (Pal & Chakravarty, 2020; Yuan et al., 2024), enhancing their survival and productivity under diverse ecological conditions (Ma et al., 2014). Whole-genome sequencing (WGS) studies have extensively explored the genetic architecture and positive selection driving economically significant traits in local cattle. These investigations have identified potential candidate genes enabling native cattle to adapt to a range of climates, from arid deserts to frigid Arctic regions (Tian et al., 2023; Yan et al., 2022a). Examples of these adaptations include heat tolerance, disease resistance, and efficient feed utilization (Ayalew et al., 2023; Chen et al., 2020; Weldenegodguad et al., 2019). For example, Mongolian cattle are well adapted to endure harsh winters (Hu et al., 2021), while Abigar cattle exhibit selection signatures in immune response genes such as *IRAK3*, *MZB1*, and *STING1*, reflecting adaptation to local disease pressures (Ayalew et al., 2023). Similarly, both Yanbian and Mongolian cattle show evidence of positive selection in *PRDM16*, a gene associated with cold stress adaptation (Yan et al., 2022). Swedish dairy cattle have developed genetic traits suited to the Nordic climate (Ghoreishifar et al., 2020; Upadhyay et al., 2019), whereas native African cattle exhibit selection for *HSPA4* and *SOD1*, genes associated with heat tolerance essential for survival in hot climates (Kim et al., 2017). Studies on Chinese cattle have confirmed selection around established functional genes such as *MC1R*, *PAG1*, and the *MYH* gene cluster, alongside the identification of novel gene candidates (Mei et al., 2018). Additionally, *SORCS3* and *SESTD1* have been linked to temperament traits in cattle (Shen et al., 2022). Such findings provide crucial insights into the development of genetic breeding strategies aimed at improving cattle productivity and adaptability. Despite the extensive body of research on cattle genetics worldwide, the genomic characteristics of local cattle breeds in Kazakhstan remain largely unexplored. Addressing this gap is necessary for uncovering the genetic potential of local cattle and leveraging this information to enhance breeding programs.

The history and culture of Kazakhstan are intrinsically linked to livestock rearing, including horses, cattle, sheep, goats, and camels, with traditional nomadic herding deeply embedded in the vast steppe regions. According to the latest data from the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan, the country maintains a cattle population of approximately 9.7 million. Among these, Kazakh white-headed and Alatau cattle are the predominant breeds used for meat and dairy production. Kazakh white-headed cattle, established in the latter half of the 20th century, are the result of crossbreeding local Kazakh and Kalmyk cattle with Hereford bulls (Kineev & Erdenov, 2005). This breed combines the exceptional meat quality of Hereford cattle with the disease resistance and adaptability of Kalmyk cattle (Kinispai et al., 2023). Notably, Kazakh white-headed cattle thrive in the harsh continental climate of the dry steppes and semi-deserts due to their resilience, low feed requirements, and ability to recover quickly after severe winters (Kinispai et al., 2023; Makaev et al., 2015). Their physiological adaptations, including well-developed skin, dense hair, and substantial subcutaneous and visceral fat layers, enable them to remain outdoors year-round without shelter (Grosheva & Levykin, 2023). Alatau cattle, a dual-purpose breed developed through crossbreeding native Kyrgyz and Kazakh cattle with Swiss and Kostroma breeds, excel in both milk and meat production (Baimukanov et al., 2021). Predominantly raised in the Almaty and East Kazakhstan regions, Alatau cattle exhibit impressive milk yields, averaging 6 300 kg annually with 3.8%–4.0% fat content, and provide high-quality meat with a slaughter rate of up to 65%. Bulls typically weigh 800–1 000 kg and gain an average of 850 grams daily under intensive fattening conditions (Baimukanov et al., 2024; Кожахметова et al., 2023). Given these characteristics, Alatau cattle are considered a valuable breed for agricultural production, with ongoing improvements expected to further enhance their productivity and efficiency.

Over the last decade, genomic studies utilizing limited microsatellite markers have identified genetic variability and candidate genes associated with economic traits and adaptability in Kazakh cattle breeds (Abdelmanova et al., 2021; Plakhtyukova & Selionova, 2021). Furthermore, single nucleotide polymorphism (SNP) analyses have revealed distinct genetic structures in Kazakh white-headed cattle compared to European and Russian breeds (Beishova et al., 2022). While these findings suggest that native Kazakh cattle breeds have undergone genetic adaptations to survive in harsh climate conditions, no studies have employed WGS to comprehensively investigate genetic variation in these breeds in this region. This study conducted the first WGS analysis of Kazakh cattle, including Kazakh white-headed cattle, Alatau cattle, and the Kazakh local breed. In addition, genomic data were combined with 272 previously published whole-genome sequences from 50 cattle populations across diverse geographic regions to assess population genetic diversity, genetic structure, and selection signatures. The findings of this study provide a foundation for future genomic research, offering new insights into the genetic architecture of Kazakh cattle. These results hold significant potential to inform breeding strategies aimed at enhancing productivity, adaptability, and resilience in the Kazakh cattle industry.

MATERIALS AND METHODS

Sample collection, DNA extraction, and sequencing

Whole-blood samples were collected from 47 cattle, including Kazakh white-headed cattle (*n*=12), Alatau cattle (*n*=12), Kazakh local breed (*n*=8), and Holstein cattle (*n*=15), with the latter sourced from a private farm in North Kazakhstan (Supplementary Table S1 and Figure S1). No mortality or adverse health effects were observed among the cattle during or after sample collection, and all animals remained in good health throughout the study.

Genomic DNA was isolated from whole-blood samples using the standard phenol-chloroform protocol (Köchl et al., 2005). Paired-end sequencing libraries with an average insert size of 318 bp were prepared for each individual, featuring an average read length of 150 bp. Sequencing was conducted using the DNBSEQ-T7 sequencing platform (MGI-Tech, China). This study was approved by the Institutional Animal Care and Use Committee of Northwest A&F University (Yangling, Shaanxi, China) and conducted in accordance with the ARRIVE guidelines.

For a global context, genomic data generated in this study were combined with whole-genome resequencing data (*n*=272) from the 1000 Bull Genomes Project, including both taurine and indicine cattle breeds (Hayes & Daetwyler, 2019) (Supplementary Table S2). This expanded dataset included a

total of 321 individuals, categorized into five groups based on breed and geographical origin (Chen et al., 2018a), including European taurine, Eurasian taurine, East Asian taurine, Indian indicine, and Chinese indicine (Supplementary Table S2). In addition, HiFi sequencing data from 17 cattle were downloaded from public databases to construct a structural variation (SV) reference set (Supplementary Table S3). Finally, two African buffalo (*Bubalus*) individuals were used as the outgroup (Supplementary Table S2).

Read mapping and SNP calling

Raw short-read data were processed using fastp (v.0.20.0) (Chen et al., 2018b) with default parameters to remove low-quality reads and trim adapter sequences. The resulting high-quality reads were aligned to the cattle reference genome (ARS-UCD1.2) using the slat module in BaseNumber (v.1.0.3), a GPU-accelerated tool for variant calling (Zhang et al., 2021), generating binary alignment MAP (BAM) files. The average sequencing depth for each individual was computed from BAM files using Qualimap (v.2.2.1) (Okonechnikov et al., 2016). Genomic variant call format (gVCF) files were generated for each sample using the slct module in BaseNumber (v.1.0.3) (Zhang et al., 2021). SNP calling was executed using the GenotypeGVCFs module in GATK (v.4.3.0.0). SNPs were subsequently filtered using BCFtools (v.1.17) (Danecek et al., 2021) based on the following criteria: sequencing depth across all individuals greater than 1/3× but less than 3× mean depth; variant confidence quality by depth (QD) greater than 2; SNPs with a root mean square (RMS) mapping quality (MQ) above 40.0; Phred-scaled probability of strand bias (FS) less than 60; Z-score from the Wilcoxon rank sum test comparing alternate and reference read mapping qualities (MQRankSum) greater than −12.5; and Z-score from the Wilcoxon rank sum test for alternate versus reference read position bias (ReadPosRankSum) greater than −8. Finally, SNPs with a minor allele count (MAC) greater than 1 were excluded to ensure the exclusion of singleton variants, thereby improving robustness of downstream analyses.

Population structure analysis

To assess population structure, SNPs were pruned using ancestry-adjusted linkage disequilibrium (LD) statistics (Bercovich et al., 2024) implemented in PCAone (v.0.4.4) with the parameters (--ld-stats 0 --ld-r2 0.8 --ld-bp 1 000 000). ADMIXTURE (v.1.3.0) (Alexander & Lange, 2011) was employed to estimate ancestral proportions, with the number of ancestral populations (*K*) ranging from 2–9. A maximum-likelihood (ML) tree was constructed using IQ-TREE (v.1.6.12) (Minh et al., 2020), with GTR+ASC as the substitution model, designating buffalo as the outgroup. The resulting phylogenetic tree was visualized using iTOL (v.6.9.1) (Letunic & Bork, 2021). Principal component analysis (PCA) was carried out using the smartPCA program in the EIGENSOFT package (v.3.0) (Patterson et al., 2006). To evaluate admixture events involving Kazakh cattle, *f*3 statistics were computed using ADMIXTOOLS (v.7.0.2) with the “qp3Pop” program (Patterson et al., 2012). In this analysis, Kazakh cattle were designated as the target population, while source populations included European taurine cattle (Norwegian Red and Kholmogory), Eurasian taurine cattle (Brown Swiss), and East Asian taurine cattle (Buriat and Hanwoo). Runs of homozygosity (ROHs) were identified using PLINK (v.1.9) (Purcell et al., 2007) with the “—homozyg” parameter, employing a sliding window of 50 SNPs (--homozyg-window-snp 50). Additional parameters were set based on previous studies (Xia et al., 2021b): --homozyg-density 50, --homozyg-snp 50, --homozyg-window-missing 5, --homozyg-window-threshold 0.05, and --homozyg-window-het 3. ROH lengths and counts were categorized into four groups: >4 Mb, 2–4 Mb, 1–2 Mb, and 0.5–1 Mb. LD decay patterns were analyzed using PopLDdecay (v.3.42), with physical distances between SNPs calculated under default parameters (Zhang et al., 2019). Nucleotide diversity ($\theta\pi$) for each breed was assessed in sliding windows of 50 kb with 25 kb increments using VCFtools (v.0.1.13) (Danecek et al., 2011).

Paternal and maternal analysis

To investigate paternal lineage, a dataset comprising 745 SNPs located within the X-degenerate region of the Y chromosome was analyzed. The dataset included samples with known paternal haplotype classifications (Chen et al., 2018a) and 153 bulls from the 1000 Bull Genomes Project (Hayes & Daetwyler, 2019). Phylogenetic relationships were inferred using the ML method implemented in IQ-TREE2 (v.1.6.12) (Minh et al., 2020) (Supplementary Table S4).

For maternal lineage analysis, mitochondrial DNA sequences were assembled using Mapping Iterative Assembler (v.1.0) (https://github.com/mpieva/mapping-iterative-assembler). These assembled sequences were combined with mitochondrial sequences with known maternal lineages (Xia et al., 2021a). Sequence alignment was performed using MUSCLE (v.5.2) (Edgar, 2022). Phylogenetic tree construction was performed using IQ-TREE2 (v.1.6.12) (Minh et al., 2020) (Supplementary Table S5).

Detection of selective sweep regions

Selective sweep regions in Kazakh cattle were identified using a multi-step approach designed to detect genomic regions under positive selection. Initially, the nucleotide diversity ($\theta\pi$) method was applied to identify selection signals within Kazakh cattle populations (Nielsen et al., 2005). $\theta\pi$ was calculated using VCFtools (v.0.1.13), applying a sliding window of 50 kb and steps of 10 kb (Danecek et al., 2011). The top 1% of windows were considered as potential selection signatures. The fixation index (F_{ST}) and cross-population extended haplotype homozygosity (XP-EHH) were used to compare Kazakh local cattle with commercial cattle breeds, including dairy breeds (Jersey and Holstein) and beef breeds (Limousin, Hereford, Angus, and Charolais). VCFtools (v.0.1.13) was employed to conduct F_{ST} analysis, using 50 kb windows with 10 kb steps (Danecek et al., 2011). Selscan (v.1.3.0) was utilized to compute XP-EHH statistics, predicting extended haplotype homozygosity for each population pair (Szpiech & Hernandez, 2014). Mean normalized XP-EHH scores were computed for each 50 kb region. Candidate regions were determined based on the top 1% of windows identified by both F_{ST} and XP-EHH. Genomic regions showing evidence of positive selection were detected through a minimum of two approaches. Candidate genes within these regions were subjected to functional Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using KOBAS (v.3.0) (Xie et al., 2011). Enrichment was determined based on Benjamini-Hochberg corrected $P<0.05$.

Detection of structural variants

To reduce false positives in SV detection caused by short

reads, an SV reference set was first constructed using 17 publicly available HiFi sequencing datasets. This reference set was generated by integrating results from three tools, including pbsv (v.2.6.2) (https://github.com/PacificBiosciences/pbsv), cuteSV (v.1.0.13) (Jiang et al., 2020), and Sniffles (v.1.0.12a) (Sedlazeck et al., 2018). HiFi reads were aligned to the reference genome using pbmm2 (v.1.12.0) (https://github.com/PacificBiosciences/pbmm2). The pbsv discover module was then employed to identify SV signatures in each BAM file, with the call module then used for SV calling. The following settings were used for HiFi data in cuteSV (v.1.0.13): "--max_cluster_bias_INS 1000 --diff_ratio_merging_INS 0.9 --max_cluster_bias_DEL 1000 --diff_ratio_merging_DEL 0.5" (Jiang et al., 2020). The parameters "-min_support 4 --report_seq --report_str --ccs_reads --genotype" were used in Sniffles (v.1.0.12a) (Sedlazeck et al., 2018). This multi-tool approach ensured robust and high-confidence SV detection across datasets. The SVs identified across individuals were merged using SURVIVOR (v.1.0.6) (Jeffares et al., 2017), consolidating data from multiple sources to create a comprehensive SV reference set. This SV reference set, derived from HiFi sequencing data, was then utilized for genotyping short-read data with Paragraph (v.2.4a) (Chen et al., 2019). To compile the final SV dataset, individual VCF files were merged using BCFtools (v.1.17). Stringent filtering criteria were applied to ensure high-quality SV calls, including: removal of SVs with a missing rate>0.1; removal of SVs with a minimum allele frequency (MAF)<0.01; and removal of SVs showing deviation from the Hardy-Weinberg equilibrium within the population (calculated using BCFtools (v.1.17) with the --hardy parameter).

RESULTS

Genome resequencing

WGS data were generated for four cattle populations in Kazakhstan: Kazakh white-headed (*n*=12), Alatau (*n*=12), Kazakh local breed (*n*=8), and Holstein cattle (*n*=15). To provide a global comparative framework, these data were integrated with available whole-genome resequencing data from the 1000 Bull Genomes Project, encompassing 272 individuals representing both humped indicine cattle (*Bos indicus*) and humpless taurine cattle (*Bos taurus*). The combined dataset included individuals assigned to 12 geographical regions based on their origins: North Europe, East Europe, Kazakhstan, Central Asia, West Europe, Central-South Europe, North-East Asia, Xizang China, North-Central China, Africa, India-Pakistan, and South-China. Moreover, two African buffalo individuals were included as an outgroup (Supplementary Tables S1, S2).

After stringent quality control, a total of 53 518 584 high-confidence SNPs were identified across the 321 samples. Among the Kazakh populations, Alatau cattle contained 14 920 214 SNPs, Kazakh local breed cattle contained 12 789 952 SNPs, and Kazakh white-headed cattle contained 15 086 508 SNPs (Supplementary Table S6). SV detection revealed a total of 192 585 bi-allelic SVs, including 87 935 deletions and 104 650 insertions (Supplementary Table S7).

Genetic structure and interactions within populations

To examine interbreed relationships among Kazakh local cattle, Alatau cattle, and Kazakh white-headed cattle, representative breeds were selected for genetic structure analysis. PCA revealed distinct genetic patterns, with samples from each geographic region clustered together. The Kazakh breeds clustered within the same genetic branch as common European cattle. As expected, Holstein cattle from Kazakhstan grouped closely with Holstein samples from the 1000 Bull Genomes Project, forming a single branch. Kazakh white-headed cattle demonstrated a closer genetic relationship to Hereford cattle, clustering within the same clade (Figure 1A; Supplementary Figure S2). Phylogenetic trees constructed using the neighbor-joining (NJ) method corroborated these findings, showing similar population affinities (Figure 1B). Both Kazakh local and Alatau cattle clustered with Brown Swiss cattle. These relationships were further validated by ADMIXTURE analysis (Supplementary Figures S3, S4). At *K*=2, all cattle breeds were divided into two distinct groups, corresponding to *Bos taurus* and *Bos indicus* (Figure 1C). At *K*=5, the genetic heterogeneity within Alatau, Kazakh local, and Kazakh white-headed cattle became evident, sharing genomic ancestry with European taurine (Hereford, Angus), Eurasian taurine (Brown Swiss, Limousin, Simmental), and East Asian taurine (Northeast Asia and Xizang China) (Figure 1C), findings that were consistent with paternal and maternal haplotype analyses (Supplementary Figures S5, S6). Notably, the genetic influence of Eurasian cattle was higher in the Alatau (65.7%) and Kazakh local breeds (64.2%) compared to European cattle (North, East, and West), whereas European taurine cattle contributed a larger proportion to the genetic makeup of Kazakh white-headed cattle (88.3%). These results indicate that the cattle population of Kazakhstan is primarily derived from three distinct ancestral lineages, further supported by *f*3 statistics (Supplementary Table S8).

Genome variation patterns

To evaluate genomic inbreeding levels in Kazakh cattle, the ROHs were calculated for each taurine cattle population. ROHs were classified based on their lengths: 0.5–1 Mb, 1–2 Mb, 2–4 Mb, and >4 Mb. Across all cattle breeds, most observed ROHs fell within the 0.5−1 Mb range. European commercial breeds exhibited a higher frequency of medium-sized (2–4 Mb) and long ROHs (>4 Mb) compared to other populations. In contrast, the total length of ROHs in the Alatau, Kazakh local, and Kazakh white-headed breeds was shorter than in European breeds but slightly higher than in African indicine cattle. Among the Kazakh populations, Alatau cattle displayed the lowest number of ROHs >4 Mb (Figure 2A; Supplementary Tables S9, S10). The average ROH-based inbreeding coefficient (FROH) for Kazakh cattle breeds ranged from 0.05 to 0.15, reflecting low levels of inbreeding and high genetic diversity (Figure 2B). In comparison, LD analysis revealed the lowest LD levels in African indicine cattle, while the highest LD was observed in Jersey cattle, followed by Charolais and Limousin breeds (Figure 2C). Kazakh local, Kazakh white-headed, and Alatau cattle exhibited faster LD decay rates than most European breeds (Figure 2C). Nucleotide diversity (θ_π) analysis further highlighted the genetic variation among cattle populations. African indicine cattle exhibited the highest θ_π, while Jersey cattle had the lowest. Among Kazakh cattle, Kazakh local (1.62×10^{-3}), Alatau (1.64×10^{-3}), and Kazakh white-headed (1.55×10^{-3}) cattle showed slightly higher θ_π values than other European breeds (Figure 2D; Supplementary Table S11).

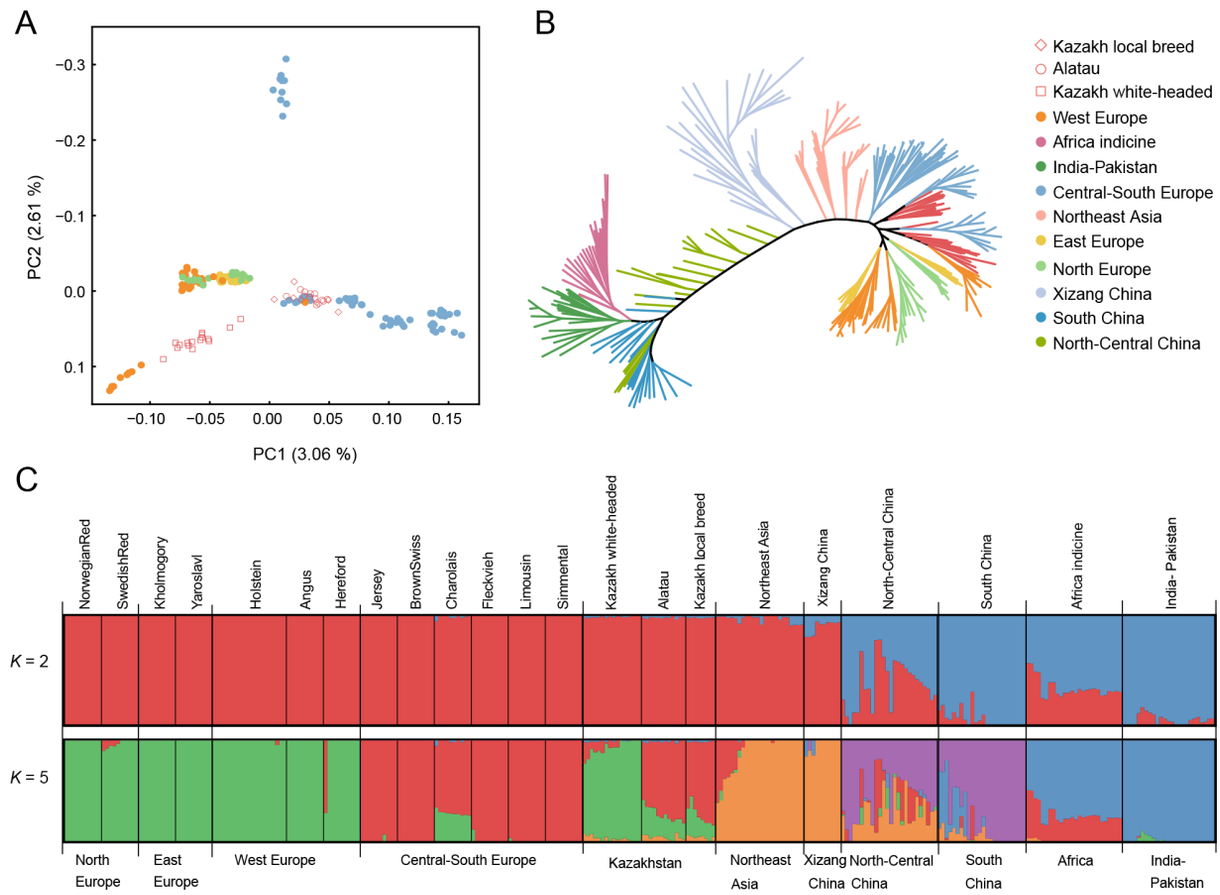

Figure 1 Phylogenetic relationship among cattle populations

A: Principal component analysis (PCA) of taurine cattle from different geographical regions. B: Neighbor-joining phylogenetic tree constructed using whole-genome SNP data. Bar represents pairwise distances between different individuals. Colors represent geographic sampling regions. C: Model-based clustering of cattle breeds using ADMIXTURE with *K*=2 and *K*=5. Colors reflect different cattle ancestries.

Genome-wide selective sweep analysis in Kazak white-headed cattle

Genomic regions under selection in Kazakh white-headed cattle were identified using three different statistical tests, F_{ST}, θ_{π}, and XP-EHH. These analyses compared Kazakh white-headed cattle and six commercial taurine cattle breeds (Jersey, Holstein, Limousin, Hereford, Angus, and Charolais). Candidate genes located within the top 1% of windows were annotated using ANNOVAR with the cattle reference genome (ARS-UCD1.2). In total, 786, 864, and 621 genes were identified through F_{ST}, θ_{π}, and XP-EHH analyses, respectively (Figure 3A; Supplementary Tables S12–S15). Of these, 77 candidate genes were shared across all three methods (Figure 3B). The most significant selective sweep region was distributed on chromosome 6 (BTA6, F_{ST}=0.6, θ_{π}=3.33, and XP-EHH=4.19). This region included a cluster of tyrosine kinase receptor genes, such as v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*), kinase insert domain receptor (*KDR*), and platelet-derived growth factor receptor alpha polypeptide (*PDGFRA*), which collectively formed a haplotype associated with strong selective pressure. These genes were clustered within a genomic region spanning 69.723 Mb to 70.613 Mb, highlighting their tight spatial association on BTA6. Haplotype analysis revealed a high degree of similarity between Hereford and Kazakh white-headed cattle (Figure 3C). The *KIT* gene, in particular, is known to play a crucial role in melanogenesis, guiding melanocyte migration from the neural crest to the skin (Besmer et al., 1993). This evidence strongly suggests that *KIT* is a key candidate gene associated with the spotting locus, which is linked to distinctive coat color patterns in Kazakh white-headed cattle.

SVs in the region of interest were analyzed to identify potential associations with coat color traits. Using HiFi sequencing data, a comprehensive set of SVs was constructed (Supplementary Table S3). This reference set was then utilized for genotyping using short reads (Supplementary Table S16), resulting in the identification of a 6 kb insertion at chromosome position 70 052 679, located upstream of the *KIT* gene. The population-level frequency of this insertion was examined across various cattle breeds (Supplementary Table S17), revealing a strong correlation with coat color distribution. Breeds characterized by white spotting on the head, such as Hereford, Kazakh white-headed, and Holstein, consistently exhibited an SV frequency of 0 for this specific insertion. Similarly, Simmental and Jersey cattle, both featuring white spots, demonstrated low frequencies of the insertion. In contrast, cattle breeds without white spotting showed significantly higher insertion frequencies, often exceeding 0.8 (Figure 3E; Supplementary Figure S7 and Table S17).

KEGG pathway and GO functional enrichment analyses were conducted to investigate the biological roles of overlapping genes identified in Kazakh white-headed cattle. Results revealed significant enrichment in 37 functional categories (corrected $P \leq 0.05$). Notably, the enriched GO

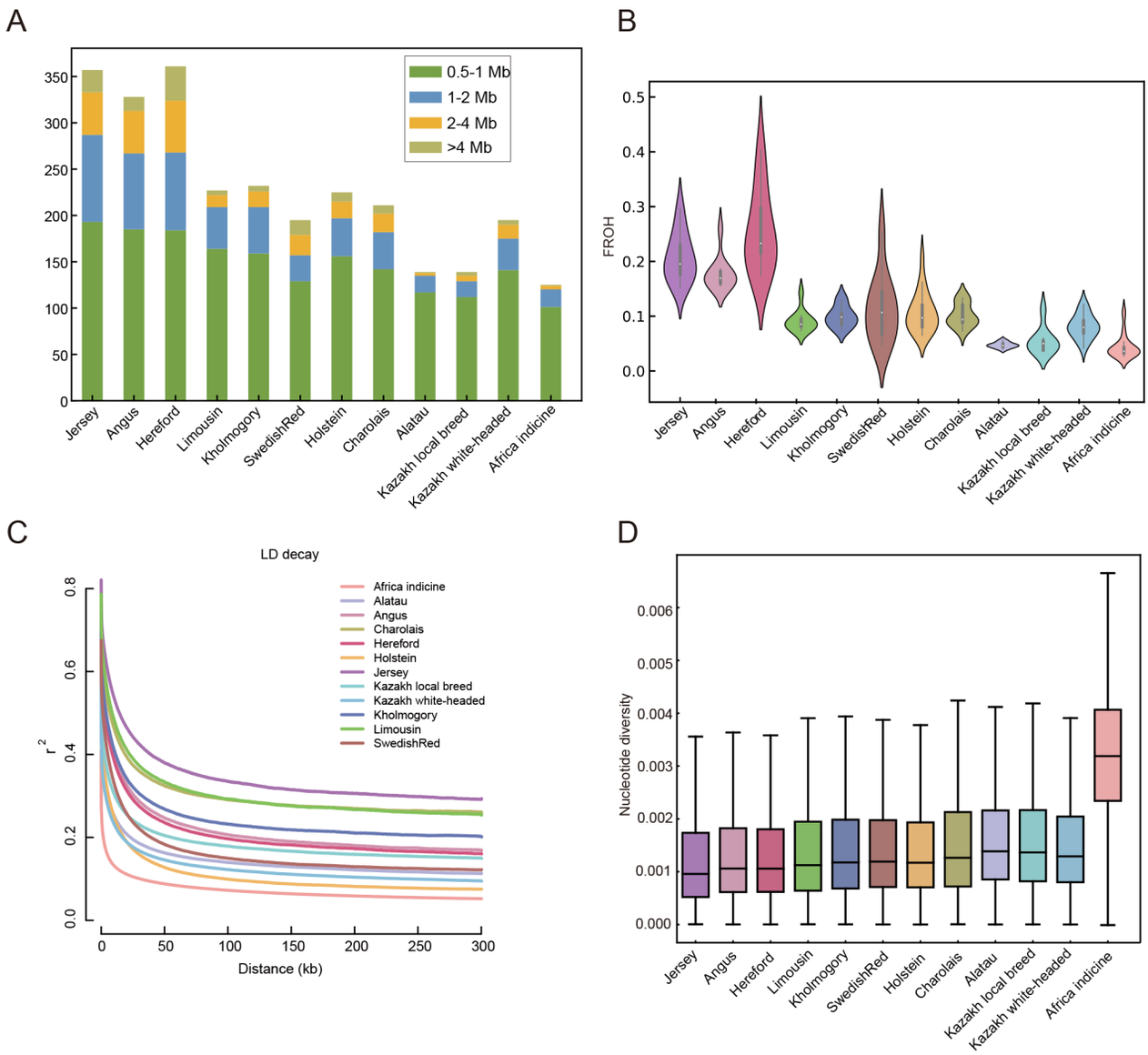

Figure 2 Summary of genomic variation statistics

A: Number of run of homozygosity (ROH) events across 12 cattle breeds. B: Inbreeding coefficients (F_i) calculated for various cattle populations. C: Linkage disequilibrium (LD) decay variation with genomic distance, quantified by correlation coefficient (r^2), in domestic cattle populations. D: Nucleotide diversity ($\theta\pi$) measured across different cattle populations.

categories included processes linked to transmembrane receptor protein tyrosine kinase activity, positive regulation of kinase activity, transmembrane receptor protein tyrosine kinase signaling pathway, peptidyl-tyrosine phosphorylation, receptor complex, and integral component of plasma membrane, melanocyte differentiation, and hematopoietic progenitor cell differentiation, suggesting a potential connection to coat color phenotypes in Kazakh white-headed cattle. Further enrichment analyses highlighted pathways associated with transforming growth factor beta receptor signaling, multicellular organism development, growth factor binding, and ATP binding, which may contribute to growth-related traits specific to Kazakh white-headed cattle (Figure 3F; Supplementary Table S18).

Scanning for highly differentiated genomic regions in Kazakh local breed and Alatau cattle

A comparative genomic analysis between the Kazakh local breed and six commercial cattle breeds (Jersey, Holstein, Limousin, Hereford, Angus, and Charolais) identified 859 genes through F_{ST} analysis and 785 genes through XP-EHH analysis (Supplementary Tables S19–S21). The overlap of these two approaches yielded 90 candidate genes (Figure 4A), associated with immune traits or diseases in the Kazakh local breed based on KEGG pathway and GO enrichment analyses. Notably, analysis revealed significant enrichment in the “Cell adhesion” category (GO:0007155) related to the immune system. Cell adhesion molecules play a crucial role in regulating and coordinating immune responses, maintaining the body’s immune balance through tissue and cellular immunity (Harjunpää et al., 2019). Cell adhesion molecules also facilitate the communication and positioning of immune cells in lymphoid tissues, promoting the activation and proliferation of lymphocytes. KEGG pathway analysis further highlighted enrichment in metabolic pathways, including “Regulation of lipolysis in adipocytes” and “cGMP-PKG signaling pathway” (Supplementary Table S22), in three key genes, *IRS1*, *PRKG1*, and *ADCY8*, which are associated with lipid metabolism and contribute to productivity traits

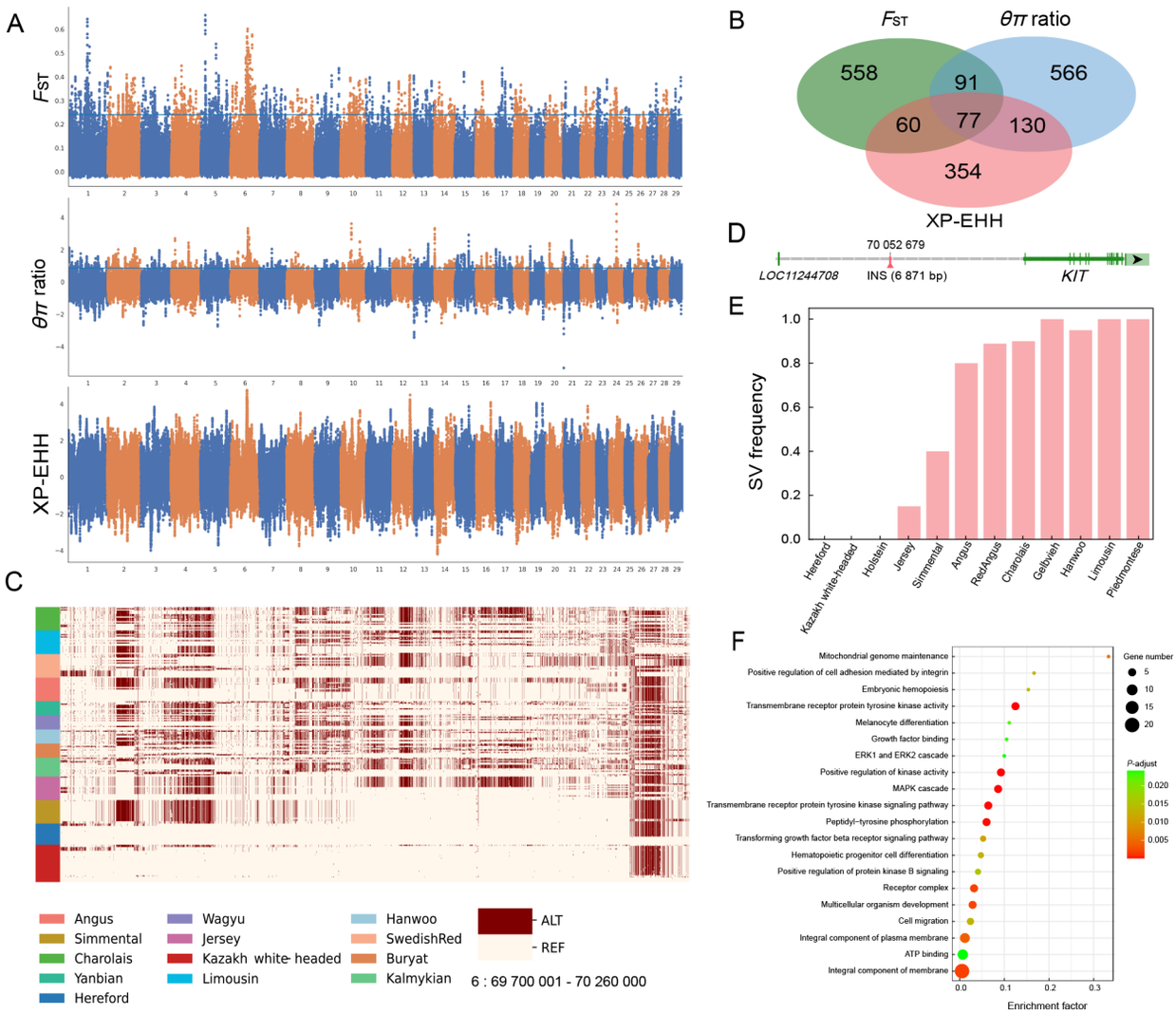

Figure 3 Analysis of positive selection signatures in Kazakh white-headed cattle genome

A: Manhattan plots depicting selective sweeps in Kazakh white-headed cattle. B: Haplotype patterns constructed from SNPs with minor allele frequencies >0.05 in the Chr 6: 69.72–70.62 Mb region. C: Venn diagram illustrating overlap of genes. D: Frequency distribution of structural variants (SVs) derived from HiFi sequencing. E: Identification of 6 kb insertion upstream of the *KIT* gene within the Chr 6: 69.72–70.62 Mb interval, based on SV data from HiFi sequencing. F: KEGG and GO pathway enrichment analyses of overlapping genes in Kazakh white-headed cattle, with pathways considered significant at corrected $P \leq 0.05$.

(Figure 4B; Supplementary Table S22).

Alatau cattle are a dual-purpose breed suitable for both milk and meat production. Comparative genomic analysis with other dairy cattle breeds was conducted to identify positive selection signatures associated with traits relevant to meat production. Using F_{ST} and XP-EHH methods, 903 and 803 selected genomic regions were identified, respectively (Supplementary Tables S23–S25). Among these, 108 candidate genes were detected by both methods in Alatau cattle (Figure 4C). To pinpoint genes linked to meat tenderness, candidate genes were subjected to GO enrichment analysis. This analysis identified *KCNMA1*, *HIF1A*, and *ANTXR1* as key genes positively influencing growth factors. Enriched GO categories included “Inner ear auditory receptor cell differentiation” (GO:0042491), “Cerebral cortex development” (GO:0021987), “Actin cytoskeleton reorganization” (GO:0031532), and “Hematopoietic progenitor cell differentiation” (GO:0002244), which are associated with beef production traits (Figure 4D; Supplementary Table S26). In addition to meat production, potential genes linked to milk production traits were identified, including *ATP2B1*, *DHX15*, *FUK*, *NEGR1*, *CCDC91*, *COG4*, and *PTK2B*. GO enrichment analysis revealed that these genes were enriched in several pathways, including “Multicellular organism development” (GO:0007275), “Protein transport” (GO:0015031), “ATP binding” (GO:0005524), “Digestive tract development” (GO:0048565), and “Feeding behavior” (GO:0007631) (Supplementary Table S27).

Comparison of the genomic regions under selection in Kazakh cattle with the expression quantitative trait locus (eQTL) database from the cattle GTEx project (Liu et al., 2022) identified 193 eGenes —genes with significant cis-eQTLs —within the selected regions (Supplementary Tables S28–S30). Among the identified genes, *FRYL*, which showed significant enrichment in the cAMP and Rap1 signaling pathways, is known to play key roles in animal growth and body size regulation (Liu et al., 2024). The down-regulation of *MYBPH* is associated with meat quality in Kazakh cattle (Yan et al., 2021). Furthermore, *S100A8*, regulated by estrogen in oviduct mucosal epithelial cells and known to influence

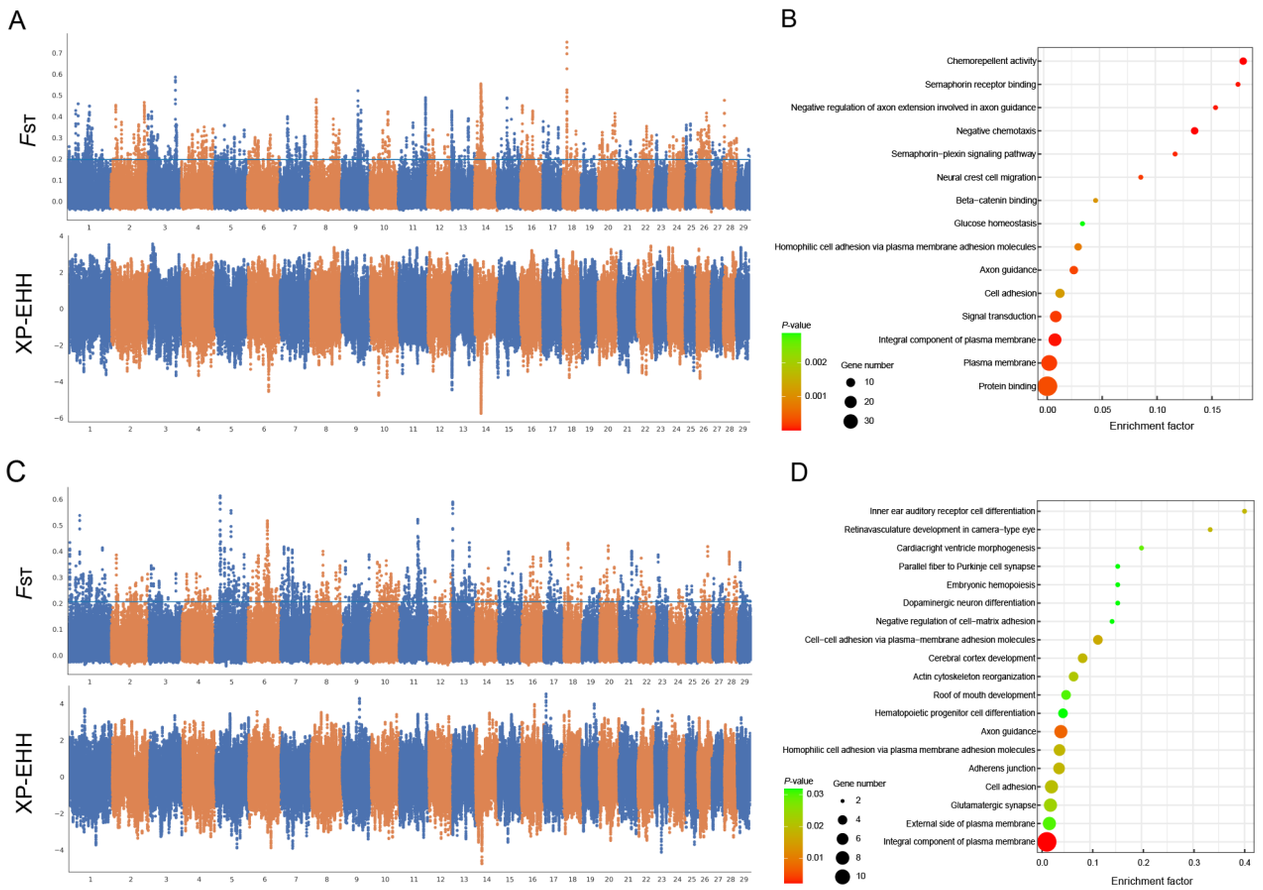

Figure 4 Analysis of selection signatures of Kazakh local breed and Alatau cattle

A, C: Manhattan plots of positive selection signatures in Kazakh local breed (A) and Alatau cattle (C) using F_{ST} and XP-EHH values. B, D: KEGG and GO pathways enriched in candidate genes in Kazakh local breed (B) and Alatau cattle (D) (corrected $P \leq 0.05$).

mucosal immune homeostasis (Li et al., 2021), is associated with a higher incidence of mastitis in cattle (Kumar et al., 2020). *PPP3R1*, which encodes a phosphatase protein associated with muscle deposition in Nelore cattle (Santos Silva et al., 2020), regulates muscle fiber apoptosis and protein degradation, leading to an increase in the myofibrillar fragmentation index (MFI) and enhanced muscle tenderness (Luo et al., 2022).

DISCUSSION

This study conducted a comprehensive WGS analysis of Kazakh cattle, shedding light on their genetic origins and the complex history of domestication in Kazakhstan. Strategically located at the heart of the Eurasian continent, Kazakhstan has historically served as a nexus between Eastern and Western civilizations. The genetic structure of Kazakh cattle reflects this geographic and cultural position, exhibiting a mix of Eastern and Western taurine ancestries. This pattern, also observed in human populations of the region (Pan et al., 2023; Seidualy et al., 2020), is likely the result of human migrations and interactions across Eurasia. Further analysis revealed that Kazakh cattle were primarily composed of three major genetic components: European, Eurasian, and East Asian taurine lineages. While East Asian taurine cattle initially spread across East Asia, this lineage was largely replaced by Eurasian taurine cattle (Chen et al., 2018a). Nevertheless, our results showed that modern Kazakh cattle retained a portion of East Asian taurine ancestry, likely representing one of the earliest taurine cattle lineages to reach Kazakhstan (Figure 1A). Despite its reduced presence due to subsequent genetic replacement, this lineage provides evidence of early cattle introductions to the region. Notably, Eurasian taurine ancestry predominated in both the Kazakh local breeds and Alatau breeds, while the European taurine lineage was more prevalent in Kazakh white-headed cattle, consistent with their breeding history, which originated through the crossbreeding of local Kazakh and Kalmyk cows with Hereford bulls (Kineev & Erdenov, 2005).

Nucleotide diversity analysis revealed that the Kazakh local, Alatau, and Kazakh white-headed cattle exhibited significantly higher diversity compared to other European breeds (Figure 2C). This increased diversity likely reflects the mixed genetic composition of Kazakh cattle and the less intensive artificial selection pressures they have experienced compared to their European counterparts. ROH analysis further supported this observation, with all three Kazakh breeds exhibiting relatively lower ROH levels compared to European breeds (Figure 2B). The higher prevalence of ROHs in European commercial breeds aligns with their history of prolonged artificial selection (Ceballos et al., 2018). Interestingly, Kazakh white-headed cattle demonstrated a higher inbreeding coefficient and lower genetic diversity compared to other Kazakh cattle, suggesting this breed may have been subjected to more intense artificial selection.

A total of 77 selected genes were identified in Kazakh white-headed cattle by combining results from F_{ST}, θ_{π}, and XP-EHH analyses. Among these, the strongest genomic signal was located on BTA6, including *KIT*, *KDR*, and *PDGFRA*, genes implicated in pigmentation and coat color phenotypes.

Haplotype analysis of these regions revealed similarities between Hereford and Kazakh white-headed cattle (Supplementary Figure S8).

The *KIT* gene is widely recognized for its role in pigmentation in mammals including cattle, horses, cats, and pigs (Cooper et al., 2006; Haase et al., 2007; Liu et al., 2009; Pielberg et al., 2002; Reinsch et al., 1999) In cattle, the vitiligo-like white spotting phenotype arises from the abnormal migration of melanocyte precursor cells from the neural crest to different skin regions (Jivanji et al., 2019). Consistent with these findings, our findings support *KIT* as a strong candidate gene responsible for the white-headed phenotype in Kazakh white-headed cattle. Additionally, *PDGFRA* was identified as another gene influencing pigmentation and melanocyte migration, affecting melanin production via the MAPK signaling pathway and *PKC* gene expression (Bennett & Lamoreux, 2003).

Structural variation is an important component of genetic diversity among cattle populations (Gong et al., 2022). In this study, a 6 kb insertion upstream of *KIT* was identified, potentially associated with the white spotting trait in Kazakh white-headed cattle (Figure 3D). Given the breeding history of Kazakh white-headed cattle, developed through crossbreeding local Kazakh and Kalmyk cows with Hereford bulls (Kineev & Erdenov, 2005), we speculate that the *KIT* haplotype originates from Hereford cattle, consistent with the phylogenetic relationships observed in our local tree analysis (Supplementary Table S8). While *KIT* appears to influence the white-headed phenotype in Kazakh white-headed cattle, additional theoretical and experimental validation is required to confirm its role.

Local cattle breeds are often characterized by robust disease resistance, a crucial adaptation for survival in challenging environments (Ma et al., 2014). Comparative analysis between the Kazakh local breed and commercial cattle breeds identified candidate genes associated with immune system functions (Figure 4B; Supplementary Table S22). In addition, several genomic regions in the Kazakh local breed were enriched in pathways such as “Regulation of lipolysis in adipocytes” and “cGMP-PKG signaling pathway”, involving *IRS1*, *PRKG1*, and *ADCY8*, key genes related to fat metabolism. Previous studies have demonstrated that *IRS1* plays an important role in the regulation of milk fat metabolism (Jiao et al., 2020), while genome-wide association studies (GWAS) have linked *PRKG1* to variations in milk fatty acid composition (Shi et al., 2019). Similarly, *ADCY8* has been implicated in lactation traits in Xinjiang Kazakh dairy horses (Liu et al., 2018). Given their established roles, we deduced that the detected genes may have an important regulatory function in fatty acid metabolism in the Kazakh local breed.

Alatau cattle are recognized as a dual-purpose breed for both dairy and beef production. Comparative genomic analyses between Alatau cattle and dairy breeds identified *KCNMA1*, *HIF1A*, and *ANTXR1* as putative candidate genes under selection, revealed through F_{ST} and XP-EHH methods. *KCNMA1* is associated with muscle growth and reproductive performance in buffaloes (De Araujo Neto et al., 2020; Li et al., 2018). *HIF1A* contributes to cellular adaptation under hypoxic conditions by mediating responses to growth factors (Lum et al., 2007; Taylor & Scholz, 2022), while *ANTXR1* regulates actin cytoskeleton reorganization pathways (Cheng et al., 2019), which can impact skeletal muscle development and growth, essential for high-quality meat production (Taye et al., 2017). In contrast, when comparing Alatau cattle to beef breeds, several genes linked to dairy production traits were identified, including *ATP2B1*, *DHX15*, *NEGR1*, *CCDC91*, *PTK2B*, *COG4*, and *FUK*. *ATP2B1* has been linked to fertility (Rahman et al., 2023), while *DHX15* has been implicated in milk protein variability (Sanchez et al., 2017). *NEGR1* is known to impact somatic scores, a widely used indicator of mastitis and udder inflammation (Mohammadi et al., 2022), while *CCDC91* is associated with fat percentage in milk (Fang et al., 2014) and *PTK2B* is associated with traits related to mastitis in dairy cows (Yang et al., 2018). The functional roles of *COG4* and *FUK* further emphasize their importance in dairy traits, particularly in protein metabolism, transport, and post-Golgi modification processes (Zhou et al., 2019). Their strong selection signals in Alatau cattle suggest they may play a pivotal role in enhancing milk protein transport and overall dairy performance.

This study offers novel insights into the genetic diversity and selection signals of Kazakh cattle, marking the first application of WGS to investigate their genetic architecture and ancestral composition. Our results provide a strong foundation for further investigation into the genetic landscape of the Kazakh cattle population, with a particular focus on traits of economic importance. This research enhances our understanding of genetic diversity, aiding in the development of breeding programs to improve productivity and resilience.

DATA AVAILABILITY

The raw sequencing data are available from the SRA database under BioProjectID PRJNA1123388, as well as the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa) under accession number PRJCA032821 and Science Data Bank (https://www.scidb.cn/en; doi: 10.57760/sciencedb.j00139.00136).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS’ CONTRIBUTIONS

Z.N. conducted comprehensive data analysis, interpreted the results, and prepared the first draft of the manuscript. Y.X., K.K., and Y.L. assisted in data analysis, validated data accuracy, and contributed to data visualization. Y.U., S.A., I.P. and N.R. were responsible for collecting and initially processing the samples. H.A.N. reviewed and edited the manuscript, suggesting substantial revisions to improve clarity and coherence. C.Y.G., M.Q., H.P.W., Y.J., and Y.D.C. designed the research framework, secured funding, supervised the entire project, and ensured coordination among all team members. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank the high-performance computing platform of Northwest A&F University and the Computing Center in Xi’an for providing computing resources.

REFERENCES

Кожахметова АН, Абуғалиев СҚ, Бупебаева ЛК. 2023. Actual issues of herd exclusion of dairy and dairy-meat cows. *Ғылым және білім*, **2**(3): 191–199. (in Russian)

Abdelmanova AS, Kharzinova VR, Volkova VV, et al. 2021. 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*, **13**(8): 351.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, **12**: 246.

Ayalew W, Wu XY, Tarekegn GM, et al. 2023. Whole-genome resequencing reveals selection signatures of Abigar cattle for local adaptation. *Animals (Basel)*, **13**(20): 3269.

Baimukanov AD, Bissembayev AT, Yuldashbayev YA, et al. 2024. Reproductive indicators of the Alatau cattle breed of Kazakhstan population. *OnLine Journal of Biological Sciences*, **24**(1): 64–70.

Baimukanov AD, Yuldashbayev YA, Demin VA, et al. 2021. Efficient breeding in Kazakhstan Alatau cattle breed population. *American Journal of Animal and Veterinary Sciences*, **16**(4): 318–326.

Beishova IS, Dossybayev KZ, Shamshidin AS, et al. 2022. Population analysis and genetic structure of two Kazakh cattle breeds using 150K SNP. *HAYATI Journal of Biosciences*, **29**(3): 301–309.

Bennett DC, Lamoreux ML. 2003. The color loci of mice-a genetic century. *Pigment Cell Research*, **16**(4): 333–344.

Bercovich U, Rasmussen MS, Li ZL, et al. 2024. Measuring linkage disequilibrium and improvement of pruning and clumping in structured populations. *BioRxiv*, doi: https://doi.org/10.1101/2024.05.02.592187.

Besmer P, Manova K, Duttlinger R, et al. 1993. The kit-ligand (steel factor) and its receptor c-kit/W: pleiotropic roles in gametogenesis and melanogenesis. *Development Supplement*: 125–137.

Ceballos FC, Joshi PK, Clark DW, et al. 2018. Runs of homozygosity: windows into population history and trait architecture. *Nature reviews Genetics*, **19**(4): 220–234.

Chen NB, Cai YD, Chen QM, et al. 2018a. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications*, **9**(1): 2337.

Chen QM, Zhan JX, Shen JF, et al. 2020. Whole-genome resequencing reveals diversity, global and local ancestry proportions in Yunling cattle. *Journal of Animal Breeding and Genetics*, **137**(6): 641–650.

Chen S, Krusche P, Dolzhenko E, et al. 2019. Paragraph: a graph-based structural variant genotyper for short-read sequence data. *Genome Biology*, **20**(1): 291.

Chen SF, Zhou YQ, Chen YR, et al. 2018b. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884–i890.

Cheng BX, Liu YZ, Zhao Y, et al. 2019. The role of anthrax toxin protein receptor 1 as a new mechanosensor molecule and its mechanotransduction in BMSCs under hydrostatic pressure. *Scientific Reports*, **9**(1): 12642.

Cooper MP, Fretwell N, Bailey SJ, et al. 2006. *White spotting* in the domestic cat (*Felis catus*) maps near *KIT* on feline chromosome B1. *Animal Genetics*, **37**(2): 163–165.

Danecek P, Auton A, Abecasis G, et al. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**(15): 2156–2158.

Danecek P, Bonfield JK, Liddle J, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, **10**(2): giab008.

De Araujo Neto FR, de Abreu Santos DJ, Fernandes Júnior GA, et al. 2020. Genome-wide association studies for growth traits in buffaloes using the single step genomic BLUP. *Journal of Applied Genetics*, **61**(1): 113–115.

Edgar RC. 2022. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications*, **13**(1): 6968.

Fang M, Fu WX, Jiang D, et al. 2014. A multiple-SNP approach for genome-wide association study of milk production traits in Chinese Holstein cattle. *PLoS One*, **9**(8): e99544.

Ghoreishifar SM, Eriksson S, Johansson AM, et al. 2020. Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. *Genetics Selection Evolution*, **52**(1): 52.

Gong M, Yang P, Fang WW, et al. 2022. Building a cattle pan-genome using more de novo assemblies. *Journal of Genetics and Genomics*, **49**(9): 906–908.

Grosheva OA, Levykin SV. 2023. Kazakh white-headed breed of cattle: the contribution of its creator K. A. Akopyan to the development of zootechnical science and adaptive animal husbandry. *Agrarian Science*, **369**(4): 62–69. (in Russian)

Haase B, Brooks SA, Schlumbaum A, et al. 2007. Allelic heterogeneity at the equine *KIT* locus in dominant white (*W*) horses. *PLoS Genetics*, **3**(11): e195.

Harjunpää H, Llort Asens M, Guenther C, et al. 2019. Cell adhesion molecules and their roles and regulation in the immune and tumor microenvironment. *Frontiers in Immunology*, **10**: 1078.

Hayes BJ, Daetwyler HD. 2019. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. *Annual Review of Animal Biosciences*, **7**: 89–102.

Hu LR, Brito LF, Abbas Z, et al. 2021. Investigating the short-term effects of cold stress on metabolite responses and metabolic pathways in Inner-Mongolia Sanhe Cattle. *Animals (Basel)*, **11**(9): 2493.

Jeffares DC, Jolly C, Hoti M, et al. 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, **8**: 14061.

Jiang T, Liu YZ, Jiang Y, et al. 2020. Long-read-based human genomic structural variation detection with cuteSV. *Genome Biology*, **21**(1): 189.

Jiao PX, Yuan Y, Zhang MM, et al. 2020. PRL/microRNA-183/IRS1 pathway regulates milk fat metabolism in cow mammary epithelial cells. *Genes (Basel)*, **11**(2): 196.

Jivanji S, Worth G, Lopdell TJ, et al. 2019. Genome-wide association analysis reveals QTL and candidate mutations involved in white spotting in cattle. *Genetics Selection Evolution*, **51**(1): 62.

Kim J, Hanotte O, Mwai OA, et al. 2017. The genome landscape of indigenous African cattle. *Genome Biology*, **18**(1): 34.

Kineev MA, Erdenov BK. 2005. Cattle breeds of Kazakhstan. Research and production center for animal husbandry and veterinary medicine. Ministry of Agriculture: Ministry of Agriculture.

Kinispai D, Ahmeevich MS, Nikolai G. 2023. Evaluation of the gene pool by GH L127V and GHR F279Y polymorphisms in Kazakh white-headed cattle. *Agrarian Bulletin of the*, **227**(12): 35–41. (in Russian)

Köchl S, Niederstätter H, Parson W. 2005. DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR. *In*: Carracedo A. Forensic DNA Typing Protocols. New Jersey: Humana Totowa, 13–29.

Kumar SR, Gupta ID, Wakchaure RS, et al. 2020. SNP discovery and association of S100A8 allelic variants with incidence of clinical mastitis in Sahiwal and Karan Fries cattle. *Journal of Entomology and Zoology Studies*, **8**(4): 1073–1077.

Leroy G, Baumung R, Boettcher P, et al. 2016. Review: sustainability of crossbreeding in developing countries; definitely not like crossing a meadow…. *Animal*, **10**(2): 262–273.

Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, **49**(W1): W293–W296.

Li J, Liu JJ, Campanile G, et al. 2018. Novel insights into the genetic basis of buffalo reproductive performance. *BMC Genomics*, **19**(1): 814.

Li XD, Cao GF, Yang HX, et al. 2021. *S100A8* expression in oviduct mucosal epithelial cells is regulated by estrogen and affects mucosal immune homeostasis. *PLoS One*, **16**(11): e0260188.

Liu DH, Li X, Wang L, et al. 2024. Genome-wide association studies of body size traits in Tibetan sheep. *BMC Genomics*, **25**(1): 739.

Liu L, Harris B, Keehan M, et al. 2009. Genome scan for the degree of white spotting in dairy cattle. *Animal Genetics*, **40**(6): 975–977.

Liu LL, Fang C, Liu WJ. 2018. Identification on novel locus of dairy traits of Kazakh horse in Xinjiang. *Gene*, **677**: 105–110.

Liu SL, Gao YH, Canela-Xandri O, et al. 2022. A multi-tissue atlas of regulatory variants in cattle. *Nature Genetics*, **54**(9): 1438–1447.

Loftus RT, MacHugh DE, Bradley DG, et al. 1994. Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, **91**(7): 2757–2761.

Lum JJ, Bui T, Gruber M, et al. 2007. The transcription factor HIF-1α plays

a critical role in the growth factor-dependent regulation of both aerobic and anaerobic glycolysis. *Genes & Development*, **21**(9): 1037–1049.

Luo H, Yang B, Li YL, et al. 2022. Effect and mechanism of energy substances, pH and myofibril fragmentation index on beef tenderness of Qinchuan cattle during postmortem ageing. *Food Science*, **43**(11): 171–179. (in Chinese)

Ma QQ, Jiao WJ, Wang ZY, et al. 2014. Tissue specificity and species superiority of cathelicidin gene expression in Chinese indigenous Min pigs. *Livestock Science*, **161**: 36–40.

Makaev SHA, Nurzhanov BS, Fomin AV. 2015. Domestic meat breed —Kazakh white-headed. *Bulletin Hearld of Beef cattle Breeding*, **4**: 57–62.

Mei CG, Wang HC, Liao QJ, et al. 2018. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Molecular Biology and Evolution*, **35**(3): 688–699.

Minh BQ, Schmidt HA, Chernomor O, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**(5): 1530–1534.

Mohammadi H, Farahani AHK, Moradi MH, et al. 2022. Weighted single-step genome-wide association study uncovers known and novel candidate genomic regions for milk production traits and somatic cell score in Valle del Belice dairy sheep. *Animals (Basel)*, **12**(9): 1155.

Nielsen R, Williamson S, Kim Y, et al. 2005. Genomic scans for selective sweeps using SNP data. *Genome Research*, **15**(11): 1566–1575.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, **32**(2): 292–294.

Pal A, Chakravarty A. 2020. Disease resistance for different livestock species. *In*: Pal A, Chakravarty AK. Genetics and Breeding for Disease Resistance of Livestock. London: Academic Press, 271–296.

Pan YW, Wen J, Ning ZL, et al. 2023. Comparative genomic and transcriptomic analyses reveal the impacts of genetic admixture in Kazaks, Uyghurs, and Huis. *Molecular Biology and Evolution*, **40**(3): msad054.

Patterson N, Moorjani P, Luo YT, et al. 2012. Ancient admixture in human history. *Genetics*, **192**(3): 1065–1093.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genetics*, **2**(12): e190.

Pielberg G, Olsson C, Syvänen AC, et al. 2002. Unexpectedly high allelic diversity at the *KIT* locus causing dominant white color in the domestic pig. *Genetics*, **160**(1): 305–311.

Plakhtyukova V, Selionova M. 2021. Influence of CAPN1 and GH genotypes on meat productivity indicators of Kazakh white-headed cattle. *In*: Muratov A, Ignateva S. Fundamental and Applied Scientific Research in the Development of Agriculture in the Far East. Cham: Springer, 121–131.

Purcell S, Neale B, Todd-Brown K, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**(3): 559–575.

Rahman JU, Kumar D, Singh SP, et al. 2023. Genome-wide identification and annotation of SNPs and their mapping in candidate genes related to milk production and fertility traits in Badri cattle. *Tropical Animal Health and Production*, **55**(2): 117.

Reinsch N, Thomsen H, Xu N, et al. 1999. A QTL for the degree of spotting in cattle shows synteny with the *KIT* locus on chromosome 6. *Journal of Heredity*, **90**(6): 629–634.

Sanchez MP, Govignon-Gion A, Croiseau P, et al. 2017. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. *Genetics Selection Evolution*, **49**(1): 68.

Santos Silva DBD, Fonseca LFS, Magalhães AFB, et al. 2020. Transcriptome profiling of muscle in Nelore cattle phenotypically divergent for the ribeye muscle area. *Genomics*, **112**(2): 1257–1263.

Sedlazeck FJ, Rescheneder P, Smolka M, et al. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods*, **15**(6): 461–468.

Seidualy M, Blazyte A, Jeon S, et al. 2020. Decoding a highly mixed Kazakh genome. *Human Genetics*, **139**(5): 557–568.

Shen JF, Chen QM, Zhang FW, et al. 2022. Genome-wide association study identifies quantitative trait loci affecting cattle temperament. *Zoological Research*, **43**(1): 14–25.

Shi LJ, Lv XQ, Liu L, et al. 2019. A post-GWAS confirming effects of *PRKG1* gene on milk fatty acids in a Chinese Holstein dairy population. *BMC Genetics*, **20**(1): 53.

Szpiech ZA, Hernandez RD. 2014. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, **31**(10): 2824–2827.

Taye M, Kim J, Yoon SH, et al. 2017. Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. *BMC Genetics*, **18**(1): 11.

Taylor CT, Scholz CC. 2022. The effect of HIF on metabolism and immunity. *Nature Reviews Nephrology*, **18**(9): 573–587.

Tian RG, Asadollahpour Nanaie H, Wang X, et al. 2023. Genomic adaptation to extreme climate conditions in beef cattle as a consequence of cross-breeding program. *BMC Genomics*, **24**(1): 186.

Upadhyay M, Eriksson S, Mikko S, et al. 2019. Genomic relatedness and diversity of Swedish native cattle breeds. *Genetics Selection Evolution*, **51**(1): 56.

Vigne JD, Peters J, Helmer D. 2005. First steps of animal domestication: new archaeozoological approaches. *In*: Proceedings of the 9th Conference of the International Council of Archaeozoology. Durham.

Weldenegodguad M, Popov R, Pokharel K, et al. 2019. Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. *Frontiers in Genetics*, **9**: 728.

Xia XT, Achilli A, Lenstra JA, et al. 2021a. Mitochondrial genomes from modern and ancient Turano-Mongolian cattle reveal an ancient diversity of taurine maternal lineages in East Asia. *Heredity (Edinburgh)*, **126**(6): 1000–1008.

Xia XT, Zhang SJ, Zhang HJ, et al. 2021b. Assessing genomic diversity and signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. *BMC Genomics*, **22**(1): 43.

Xie C, Mao XZ, Huang JJ, et al. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, **39**(suppl_2): W316–W322.

Yan CL, Lin J, Huang YY, et al. 2022. Population genomics reveals that natural variation in PRDM16 contributes to cold tolerance in domestic cattle. *Zoological Research*, **43**(2): 275–284.

Yan XM, Wang J, Li HB, et al. 2021. Combined transcriptome and proteome analyses reveal differences in the longissimus dorsi muscle between Kazakh cattle and Xinjiang brown cattle. *Animal Bioscience*, **34**(9): 1439–1450.

Yang F, Chen FH, Li LL, et al. 2018. GWAS using 2b-RAD sequencing identified three mastitis important SNPs via two-stage association analysis in Chinese Holstein cows. *BioRxiv*, doi: https://doi.org/10.1101/434340.

Yuan YG, Liu SZ, Farhab M, et al. 2024. Genome editing: an insight into disease resistance, production efficiency, and biomedical applications in livestock. *Functional & Integrative Genomics*, **24**(3): 81.

Zhang C, Dong SS, Xu JY, et al. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, **35**(10): 1786–1788.

Zhang Q, Liu H, Bu FX. 2021. High performance of a GPU-accelerated variant calling tool in genome data analysis. *BioRxiv*, doi: https://doi.org/10.1101/2021.12.12.472266.

Zhou CH, Li C, Cai WT, et al. 2019. Genome-wide association study for milk protein composition traits in a Chinese Holstein population using a single-step approach. *Frontiers in Genetics*, **10**: 72.