

Whole-genome sequencing of 1 363 individuals reveals population structure and local adaptation of Xinjiang indigenous sheep

Chang-Chun Lin^{1,#}, Yuan Xu^{2,#}, Cai-Yue Gao², Jun-Cheng Huang¹, Hanikzi Tulafu¹,
Peng-Fei Li¹, Ran Li^{2*}, Yu Jiang^{2*}, Wei-Wei Wu^{1*}

¹ Key Laboratory of Genetic Breeding and Reproduction for Cashmere Goat and wool sheep in Xinjiang; Key Laboratory of Evaluation and Utilization of Livestock and Poultry Resources Sheep; Xinjiang Uygur Autonomous Region Sheep and Goat Engineering Research Center; Institute of Animal Science, Xinjiang Uygur Autonomous Region Academy of Animal Sciences, Urumqi 830000, China

² Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, 712100, China

*Corresponding author. Weiwei Wu: wuweiweigp@foxmail.com; Yu Jiang: yu.jiang@nwfau.edu.cn; Ran Li: ran.li@nwfau.edu.cn

Foundation items: This work was supported by the Major Science and Technology Projects of Xinjiang Uygur Autonomous Region (Grant No. 2024A02004), the Xinjiang Research Project on Excavation and Innovative Utilization of Rare and Endangered Ruminant Livestock Germplasm Resources (Grant No. 2022TSYCCX0045), and the Autonomous Region Fleece Sheep Breed Selection and Breeding Enhancement Program (Grant No. XJARS-09-06)

Chang-Chun Lin, Yuan Xu contributed equally to this work.

ABSTRACT

Indigenous sheep breeds in Xinjiang, China, display marked phenotypic diversity across heterogeneous environments, yet their population structure and evolutionary history of key traits remain incompletely resolved. We analyzed whole-genome resequencing data from 1 363 individuals representing 14 indigenous Xinjiang sheep populations. Population genomic analyses revealed three major genetic clusters. Cele Black and Karakul sheep formed two distinct lineages that diverged from the main cluster approximately 2 000 years ago, whereas the remaining breeds separated more recently, approximately 1 000 years ago. Trait-focused selection scans identified candidate loci associated with reproductive seasonality, coat color, and fat-rumped morphology. Comparisons between year-round and seasonal breeders showed strong selection at *TSHR* in year-round estrus breeds. Fine mapping prioritized a candidate regulatory variant (g. 90612698 T>G), which was enriched in year-round breeders and associated by expression quantitative trait locus (eQTL) analysis with reduced *TSHR* expression in thyroid tissue. Coat-color comparisons identified a high-frequency *KIT* variant (g.70935444 G>A) in Cele Black and Karakul sheep, with allele-frequency trajectory inference indicating a recent rise approximately 400 years ago. For fat-rumped sheep, ancestral recombination graph reconstruction traced the genealogical history of a *TBXT*-associated variant (g.88882077 C>A), whose derived allele increased in frequency approximately 1 800 years ago. Overall, our results provide a genome-wide view of genetic diversity and demographic history in Xinjiang sheep and highlight



candidate variants underlying economically important traits, offering resources for conservation and molecular breeding.

Keywords : Whole-genome resequencing; Population structure; Genetic diversity; Demographic history; Xinjiang indigenous sheep

INTRODUCTION

Against the backdrop of accelerating global climate change and the transition of livestock production systems toward sustainability, elucidating livestock population structure and environmental adaptation has become an important research topic in evolutionary biology and precision breeding (Hayes et al., 2013; Passamonti et al., 2021). Local breeds have been jointly shaped by long-term natural selection and region-specific production systems and therefore often harbor abundant adaptive genetic variation. As a result, they represent important genetic resources for investigating environmental adaptation and breed formation in domestic animals (Hoffmann, 2010; Hoffmann, 2013; Leroy et al., 2018). Xinjiang is an important eco-geographical region in northwestern China and provides an ideal setting for studying genetic diversity, population differentiation, and trait-associated variation in local livestock. This region occupies a strategic inland position in Eurasia and historically served as a major hub along the Silk Road (Chan, 2025). In addition, Xinjiang is characterized by complex topography, diverse ecosystems, and marked environmental heterogeneity (Yao et al., 2022). Under the combined effects of variation in climate, water availability, vegetation, and local production systems, 14 indigenous sheep breeds have formed in Xinjiang and



exhibit substantial phenotypic divergence (Yao et al., 2022; Zhang et al., 2023a; Zhang et al., 2023b). Previous studies have revealed the lineage history of Chinese indigenous sheep (Zhao et al., 2017), and recent work on Xinjiang sheep has provided important insights into environmental adaptation, gene flow, and selection signatures using mitochondrial, single-nucleotide polymorphism (SNP) array, genome-environment association analyses, and whole-genome data (Zhang et al., 2023a; Li et al., 2024). Nevertheless, further analyses based on larger whole-genome datasets, broader breed representation, more detailed trait-focused comparisons, and integration of functional genomic evidence are needed to refine our understanding of the genetic structure, evolutionary history, and genetic basis of economically important traits in Xinjiang indigenous sheep.

Beyond the population structure and genetic diversity described above, the indigenous sheep breeds of Xinjiang also exhibit remarkable adaptive and phenotypic diversity, making them valuable materials for investigating the genetic basis of economically important traits in sheep (Yao et al., 2022; Zhang et al., 2023a; Zhang et al., 2023b; Li et al., 2024). For example, Duolang sheep are characterized by year-round estrus, early puberty, and relatively high reproductive efficiency, whereas Kazakh sheep display a typical fat-rumped phenotype (Zhu et al., 2016; Zhu et al., 2025; Bi et al., 2025). Across ecological settings ranging from deserts and alpine regions to oasis-grassland ecotones, Xinjiang indigenous sheep exhibit marked divergence in reproductive rhythm, including both perennial-estrous and seasonal-estrous breeds; however, the genetic basis underlying this divergence remains to be further elucidated (Zhang et al., 2023a;



Zhu et al., 2023; Bi et al., 2025). In addition, Xinjiang sheep show substantial variation in tail morphology and coat color (Zhu et al., 2016; Han et al., 2019; Yao et al., 2019). In particular, although previous studies have linked the short-tail phenotype in fat-rumped sheep to *TBXT*-related variation, the population origin and evolutionary history of fat-tail and fat-rump traits have not yet been fully resolved. Therefore, we performed whole-genome resequencing and population-genetic analyses of 1 363 individuals representing 14 indigenous Xinjiang sheep populations, generating a high-quality variant map with substantially expanded sample size and breed coverage. Using this dataset, we characterized genetic diversity, population structure, maternal and paternal lineages, and demographic history in Xinjiang indigenous sheep. We further conducted trait-focused genome-wide selection scans for reproductive rhythm, coat color, and tail morphology, and integrated haplotype patterns, protein-level predictions, Sheep genotype-tissue expression (GTEx) and eQTL evidence, and allele-frequency trajectory inference using Relate/CLUES to prioritize candidate genes and putative functional variants associated with phenotypic variation. Together, these analyses extend previous work by providing a broader and more functionally informed genomic framework for conserving Xinjiang sheep genetic resources and enabling molecular breeding.

MATERIALS AND METHODS

Ethics declarations

All animal related experimental procedures adhered to the Regulations on the Management of Laboratory Animals in China and were sanctioned by the Animal Health and Ethics Committee of the Institute of Animal Science, Xinjiang Academy of Animal Husbandry Sciences (Approval No.202304).

Sample collection, DNA extraction, and sequencing

A total of 1 363 sheep whole-genome samples representing 14 indigenous Xinjiang sheep populations were included in this study. The newly sequenced cohort comprised 1 100 individuals from 13 indigenous sheep breeds in Xinjiang, China, including Kazakh sheep (HSK, $n=58$), Altay sheep (ALTs, $n=59$), Bashbay sheep (BSB, $n=60$), Kirghiz sheep (KEKZ, $n=65$), Tashkurgan sheep (TSG, $n=60$), Duolang sheep (DL, $n=25$), Baerchuke sheep (BECK, $n=60$), Bayinbuluke sheep (BYK, $n=135$), Yecheng sheep (YC, $n=100$), Lop sheep (LB, $n=22$), Cele Black sheep (CLH, $n=170$), Turfan Black sheep (TLFH, $n=90$), and Karakul sheep (KLKE, $n=196$). All newly sequenced samples were collected from privately owned farms across Xinjiang. No mortality or adverse health effects were observed during or after sampling, and all animals remained in good health throughout the study period. In addition, 263 publicly available whole-genome sequencing datasets were incorporated, including 197 Hetian sheep (HT; CNP0007875), 16 Lop sheep (LB; PRJNA645671 and PRJCA001227), and 50 Duolang sheep (DL; PRJNA1177205). Detailed sample information is provided in Supplementary Table S1.

Approximately 5 mL of whole blood was collected from the jugular vein of each sheep into vacuum tubes containing K2EDTA as an anticoagulant. Genomic DNA was

131 extracted from blood samples using a commercial blood DNA extraction kit (Tengen
132 Bio, Beijing, China) according to the manufacturer's instructions, and DNA integrity
133 was assessed by 1% agarose gel electrophoresis. DNA purity and concentration were
134 measured with a NanoDrop 2 000 spectrophotometer (Thermo Fisher Scientific, USA).
135 Genomic DNA was sheared to approximately 150 bp using a Bioruptor Pico ultrasonic
136 sonicator. Sequencing libraries were prepared following the manufacturer's protocol
137 (MGI, China) and were subsequently sequenced on a DNBSEQ-T7 platform for whole-
138 genome resequencing.

139 **Sequence processing, read mapping and variant calling**

140 FASTQ files for publicly available datasets were generated from raw sequencing data
141 using sequence read archive (SRA) Toolkit (v.2.9.2) (Kodama et al., 2012). To
142 minimize potential batch effects and coverage-related differences between newly
143 generated and publicly available datasets, all samples were processed using the same
144 bioinformatics workflow. Publicly available raw sequencing data were converted to
145 FASTQ format and analyzed together with the newly generated data. To ensure data
146 quality for downstream analyses, raw reads were quality-filtered using fastp (v.0.19.4)
147 (Chen et al., 2018). After adapter trimming and removal of low-quality bases, clean
148 reads were aligned to the sheep reference genome (ARS-UI_Ramb_v2.0) using BWA-
149 MEM (v.0.7.17) with default parameters (Vasimuddin et al., 2019). The Samtools (v.1.2)
150 (Li et al., 2009) was used for format conversion, sorting, indexing, and basic alignment
151 processing. Polymerase chain reaction (PCR) duplicates were marked and removed



using Picard (v.2.18.17). Variants were called to generate a raw variant set, followed by stringent filtering to obtain high-confidence genotype data.

Initial SNP calling was performed with GATK HaplotypeCaller (v.4.6.1) under default parameters (McKenna et al., 2010). The merged raw variant set was filtered using GATK VariantFiltration with the following thresholds: $QUAL < 30.0$, $QD < 2.0$, $FS > 60.0$, $MQ < 40.0$, $SOR > 3.0$, $MQRankSum < -12.5$, and $ReadPosRankSum < -8.0$. To reduce the influence of sequencing-depth differences among samples, we further applied individual-level depth-based genotype masking using a custom Python script, `mask_gt_by_depth_gz.py`. For each individual, genotypes with a site depth lower than one-third of the individual mean depth or higher than three times the individual mean depth were masked as missing. SNPs with minor allele frequency (MAF) < 0.01 and missing rate > 0.2 were removed using VCFtools (v.0.1.13) (Danecek et al., 2011). All retained SNPs were annotated with ANNOVAR (Wang et al., 2010) against the sheep reference genome and classified by genomic context into exonic (synonymous and nonsynonymous), intronic, and upstream/downstream categories. High-quality SNPs were phased using Beagle (beagle.06Aug24.a91.jar) (Browning et al., 2018).

Population structural analysis

To balance sample sizes across groups, we randomly selected 50 individuals per breed (except Lop sheep) (Puechmaille, 2016; Lawson et al., 2018). Prior to PCA and ADMIXTURE analyses, linkage disequilibrium (LD) pruning was performed using PLINK (v.1.90b7) to reduce marker redundancy and minimize the influence of highly correlated dense markers on population structure inference (Purcell et al., 2007).



Specifically, SNPs were pruned using a sliding window of 50 SNPs, a step size of 10 SNPs, and an r^2 threshold of 0.05 (--indep-pairwise 50 10 0.05). This relatively stringent threshold was used to retain approximately independent markers and reduce potential artificial clustering caused by local LD, long haplotypes, or ROH-enriched regions, particularly in breeds with elevated inbreeding. Principal component analysis (PCA) was performed using EIGENSOFT (v.6.1.4) with default parameters to assess genetic structure and variation among breeds and to visualise genetic relationships among individuals (Patterson et al., 2006). Phylogenetic relationships were inferred using the neighbor-joining method (Saitou & Nei, 1987), and the resulting trees were visualized in interactive tree of life (iTOL) (Letunic & Bork, 2024). Population structure was inferred with ADMIXTURE (v.1.3.0) to estimate individual ancestry proportions across breeds (Dominguez Mantes et al., 2023). We evaluated K values from 2 to 9, performed 10 independent runs per K , and selected the K with the lowest cross-validation (CV) error as the optimal number of ancestral clusters.

Population diversity analysis

Linkage disequilibrium decay was estimated using PopLDdecay (v.3.43) with default parameters (Zhang et al., 2019). Genome-wide nucleotide diversity (π) was estimated for each breed in sliding windows using VCFtools (v.0.1.13) based on the filtered high-quality SNP dataset. For the supplementary comparison, the same pipeline and filtering criteria were applied to the additional Chinese indigenous breeds, Chinese Merino sheep, and the Iranian indigenous sheep breed. To assess genetic diversity across breeds, we estimated observed and expected heterozygosity (H_o and H_e) from the filtered high-



quality SNP dataset using VCFtools (v.0.1.13). Runs of homozygosity (ROHs) were identified in individual genomes using the --homozyg module in PLINK (v.1.90b7). ROHs were called using a sliding window of 50 SNPs (--homozyg-window-snp 50), a minimum ROH length of 1 000 kb (--homozyg-kb 1 000), a minimum SNP density of one SNP per 50-kb (--homozyg-density 50), and a maximum gap of 1 000 kb between adjacent SNPs (--homozyg-gap 1 000). Within each window, up to 3 heterozygous sites (--homozyg-window-het 3) and 5 missing genotypes (--homozyg-window-missing 5) were allowed, and at least 95% of markers were required to meet homozygosity criteria (--homozyg-window-threshold 0.05). Parameter settings followed established protocols from previous studies (Zhannur et al., 2025). Subsequently, the ROH-based genomic inbreeding coefficient (F_{ROH}) was calculated as the total length of ROHs per individual divided by the total length of the analysed autosomes. Breed-specific F_{ROH} distributions were visualised using ggplot2 in R (v.4.4.1). Violin plots overlaid with boxplots were used to summarise the distribution of F_{ROH} among individuals within each breed. Breeds were ordered from highest to lowest according to the median F_{ROH} .

To assess maternal and paternal lineages, mitochondrial FASTA sequences were generated from BAM files using Analysis of Next Generation Sequencing Data (ANGSD) (Korneliussen et al., 2014) and aligned with MUSCLE, followed by maximum-likelihood phylogenetic reconstruction using IQ-TREE for mitochondrial haplogroup assignment (Trifinopoulos et al., 2016). For Y-chromosome analysis, a high-quality sheep Y-chromosome genome sequence was used as the reference, and Y-



chromosome-specific diagnostic variants were selected to assign paternal haplotypes (Meadows et al., 2006; Deng et al., 2020).

Population history analysis

For Multiple Sequentially Markovian Coalescent 2 (MSMC2)-based inference, HT was excluded due to insufficient sequencing depth, and LB was excluded because too few individuals met the coverage criteria. From each of the remaining 12 indigenous Xinjiang sheep breeds, including HSK, ALTs, BSB, KEKZ, TSG, DL, BECK, BYK, YC, CLH, TLFH, and KLKE, we selected the two individuals with the highest sequencing depth. This strategy was used to maximize genotype reliability for MSMC2 inference while maintaining the same number of individuals across breeds. The sequencing depth information for the selected individuals is provided in the sample information Supplementary Table S2. Genome-wide mappability masks were generated using makeMappabilityMask.py (Wang et al., 2020) under the SNPable framework. Divergence times between breeds were inferred using MSMC2 (v.2.1.2) by estimating relative cross-coalescence rates for each breed pair (Schiffels & Durbin, 2014). Time segments were specified as $1 \times 2 + 25 \times 1 + 1 \times 2$. Historical effective population sizes were inferred using the estimate module in SMC++ (v.1.15.3), applying the same mappability mask as in MSMC2 to exclude low-confidence regions (Terhorst et al., 2017). We assumed a mutation rate (μ) of 1.51×10^{-8} per site per generation (Li et al., 2020) and a generation time (g) of 3 years (De Magalhães & Costa, 2009). These parameter values were used for all downstream demographic inferences. Demographic

trajectories were inferred over 1–40 000 generations using default knot settings, with effective population size modeled using a cubic spline.

Detection of selective signals associated with key phenotypic traits

To identify genomic regions under selection associated with key phenotypic differentiation in Xinjiang indigenous sheep, we applied three SNP-based statistics: nucleotide diversity ratio ($\theta\pi$ ratio), fixation index (F_{ST}) (Weir & Cockerham, 1984), and cross-population extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007). To reduce potential bias caused by unequal sample sizes among breeds, trait-focused selection-signal analyses were performed using balanced sample sets. Specifically, 50 individuals were randomly selected from each breed whenever possible; for breeds with fewer than 50 available individuals, all available individuals were retained. This balanced dataset was used for calculating F_{ST} , $\theta\pi$ ratio, and XP-EHH in phenotype-based comparisons. All statistics were computed using 50-kb sliding windows with a 25-kb step size. For each selection statistic, candidate windows were defined as those falling within the top 1% of the corresponding genome-wide empirical distribution. Overlapping windows that exceeded the corresponding top 1% cutoff in at least two of the three statistics were designated as strong candidate regions for downstream analyses. To account for multiple testing across the large number of sliding windows evaluated genome-wide, empirical significance of F_{ST} -based candidate windows was further assessed using 1 000 genome-wide permutations. For each F_{ST} window, a genome-wide empirical adjusted P value was calculated and reported as `genome_wide_emp_padj`. Candidate F_{ST} windows with `genome_wide_emp_padj`<0.05 were considered



statistically significant after genome-wide empirical correction. In contrast, genotype frequencies of candidate variants were calculated using all available individuals and are presented as within-breed or within-phenotype-group proportions rather than raw genotype counts.

To explore potential regulatory consequences of candidate genes, we quantified their expression profiles across tissues using the sheep GTEx resource (Gong et al., 2025). Empirical *P*-values were obtained using permutations tests and adjusted for multiple testing using the false discovery rate (FDR) method. eGenes were defined as genes with an $FDR < 0.05$, as implemented in eGene_detection.R. For the *TSHR* gene, we defined the lead cis-eQTL as the variant with the lowest nominal *P* value. We then calculated LD between the lead eQTL variant and nearby variants to characterize local haplotype structure. Finally, genotype-dependent effects were visualised by comparing gene expression across genotype classes at the lead eQTL.

Estimation of derived allele frequency trajectories

To infer temporal changes in derived allele frequencies, we first used KING (v.2.3.1) to identify and remove individuals related within the third degree (Manichaikul et al., 2010). We retained only breeds containing more than 40 unrelated individuals after filtering and with high mean sequencing depth. Consequently, HT were excluded because of insufficient sequencing depth. The final dataset comprised six breeds: BSB ($n=60$), CLH ($n=132$), DL ($n=45$), KLKE ($n=46$), BECK ($n=53$), and BYK ($n=121$). To assign ancestral and derived states, we used MafFilter (v.1.3.1) to process multi-species sequence alignments among sheep, cattle, goat, and takin (Dutheil et al., 2014).



We then used *est-sfs* (v.2.04) to estimate, under a probabilistic framework, the posterior probability that the major allele represents the ancestral state, thereby accounting for uncertainty in ancestral allele inference (Keightley & Jackson, 2018). We reconstructed ancestral recombination graphs (ARGs) for each breed using *Relate* (v.1.2.1), assuming a generation time of $g=3$ years and a mutation rate of $\mu=1.51\times10^{-8}$ per site per generation (Li et al., 2020), with recombination rates specified according to a physical map (Speidel et al., 2019). Finally, to characterize genealogical histories of candidate SNPs, we applied *CLUES*, which leverages local tree topologies inferred by *Relate* to model allele-frequency trajectories through time under a selection framework (Stern et al., 2019).

Bioinformatics analysis of mutant loci

Protein molecular weight, theoretical isoelectric point (pI), instability index, and hydrophilicity were computed using *ExPASy ProtParam* and *ProtScale*. Three-dimensional (3D) structures of the sheep *TBXT* and *KIT* missense variants were predicted using the *I-TASSER* server (Iterative Threading ASSEmbly Refinement), and structural models were visualised in *PyMOL*.

Functional enrichment analysis

Candidate genes identified from selection scans were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis using the *KOBAS* platform (Bu et al., 2021), and results were summarized and visualized with the R package *clusterProfiler*. Enrichment terms with adjusted *P* values <0.05 were considered statistically significant.



RESULTS

Genetic variation and annotation of SNPs

We generated 1 100 newly sequenced whole genomes from 13 indigenous Xinjiang sheep breeds, comprising TLFH ($n=90$), DL ($n=25$), TSG ($n=60$), BYK ($n=135$), ALTs ($n=59$), HSK ($n=58$), BSB ($n=60$), KEKZ ($n=65$), LB ($n=22$), BECK ($n=60$), YC ($n=100$), CLH ($n = 170$), and KLKE ($n = 196$) (Figure 1). In addition, we incorporated 263 publicly available genomes DL ($n = 50$), HT ($n=197$), LB ($n=16$), yielding a dataset of 1 363 samples for read alignment, variant calling, and variant level quality control (Supplementary Table S1). In total, the final dataset comprised 1 363 whole-genome samples from 14 indigenous Xinjiang sheep populations, including 1 100 newly sequenced individuals from 13 breeds and 263 publicly available genomes. The public samples consisted of 197 HT, 16 LB, and 50 DL. Newly sequenced and public samples are distinguished in Figure 1B. On average, 28.35 Gb of sequence data were generated per individual, yielding a population-wide mean sequencing depth of $8.36\times$ (range: $4.52\times$ – $18.87\times$). Quality-controlled reads were aligned to the reference genome (ARS-UI_Ramb_v2.0) with a mean mapping rate of 99.89% (Supplementary Table S3). After quality filtering, 35 103 309 high-confidence SNPs were retained across all samples. Of these SNPs, 82.03% were supported by records in the most recent National Center for Biotechnology Information (NCBI) sheep SNP database (last accessed: November 6, 2025). We identified 6 307 960 novel SNPs, with the largest number detected in the HSK breed (5 817 968 novel SNPs), thereby substantially expanding the catalog of

genomic variation in sheep. These findings indicate that the HSK population harbors a considerable number of previously unreported SNPs, a substantial proportion of which are shared across different breeds (Supplementary Table S3). Functional annotation indicated that 59.31% of SNPs were intergenic, whereas 40.69% mapped to genic regions, including exons, introns, and untranslated regions (UTRs) (Supplementary Table S4). Within genic regions, introns contained the largest share of SNPs (12 417 208; 35.37%). In total, 239 516 SNPs (0.68%) fell within exons, including 133 065 synonymous variants (0.30%) and 94 651 missense variants (0.27%).

Population structure and genetic diversity of indigenous sheep in Xinjiang

To assess genetic relationships among 14 indigenous sheep populations in Xinjiang, we performed population structure analyses (Figure 2). A neighbor-joining (NJ) phylogenetic tree was constructed using Asiatic Mouflon (AMF) as an outgroup. The NJ tree showed that most breeds clustered into largely distinct clades. Specifically, ALTs, BSB, HSK, and BYK from northern Xinjiang grouped into one major clade, whereas YC, TSG, KEKZ, and DL from southern Xinjiang grouped into another. This pattern is broadly consistent with geographic distribution and suggests pronounced geographic structuring. Nevertheless, some breeds did not strictly follow a geographic clustering pattern; for instance, TLFH clustered closer to several southern Xinjiang breeds, indicating possible historical gene flow or shared ancestral components. Principal component analysis broadly partitioned the 14 breeds into three clusters: CLH, KLKE, and the remaining Xinjiang breeds (Figure 2B, C). ADMIXTURE cross-validation errors across different K values are shown in Supplementary Figure S1. In



combination with the PCA and phylogenetic results, we focused on $K=3$ and $K=5$ as representative and biologically interpretable levels of population structure. At $K=3$, ADMIXTURE analysis produced a clustering pattern largely concordant with the PCA results (Figure 2D). At $K=5$, DL, TSG, and KEKZ separated from CLH and KLKE, revealing additional substructure among Xinjiang sheep breeds. Results for additional K values are provided in Supplementary Figure S2. Although the NJ tree, PCA, ADMIXTURE, and demographic analyses revealed clear population-genetic structure among Xinjiang sheep breeds, these broad genetic or geographic groupings do not necessarily correspond to ecological categories. Therefore, we did not use the northern Xinjiang, southern Xinjiang, or early-diverging lineages as ecological groups for interpreting local adaptation. Instead, each breed was further annotated according to its main ecological and landscape background.

Specifically, ALTs, BSB, and HSK were mainly associated with mountain steppe, temperate grassland, and northern Xinjiang pastoral environments; BYK was associated with high-altitude grassland and alpine meadow environments in central Xinjiang; BECK, DL, LB, YC, and TLFH were mainly associated with desert-oasis systems, basin environments, semi-desert rangelands, and oasis-pastoral systems; KEKZ, TSG, and KLKE were associated with plateau, piedmont, mountain-pasture, and high-altitude environments in southwestern Xinjiang; and CLH was associated with riparian grassland and oasis-pastoral environments. These breed-level ecological annotations were added to Supplementary Table S5 and were used to provide environmental context for interpreting candidate signals of local adaptation.



To evaluate genetic diversity across 14 Xinjiang indigenous sheep populations, we quantified linkage disequilibrium (LD), nucleotide diversity, and ROH (Figure 3). The LD analyses revealed breed-specific decay patterns: BECK showed the fastest decay, whereas KLKE exhibited the slowest decay, followed by LB. LD decay curves approached a plateau within approximately 100 kb across all breeds (Figure 3A). Nucleotide-diversity estimates were broadly comparable among the 14 Xinjiang sheep breeds, with BYK showing the highest π and LB the lowest. Across breeds, π values fell within a narrow range of 2.79×10^{-3} to 3.01×10^{-3} (Figure 3B; Supplementary Table S4). To place the genetic diversity of Xinjiang sheep in a broader context, we further compared genome-wide nucleotide diversity with other Chinese indigenous breeds, an improved Merino-derived breed, and an international local breed, including Hu sheep, Tibetan sheep, Tan sheep, Ujimqin sheep, Yunnan sheep, Chinese Merino sheep, and an Iranian indigenous sheep breed (Supplementary Figure S3). The expanded comparison showed that Xinjiang sheep exhibited moderate to relatively high nucleotide diversity, with π values largely comparable to those observed in other Chinese indigenous and international reference breeds. Observed heterozygosity (H_o) ranged from 0.21 to 0.27, whereas expected heterozygosity (H_e) ranged from 0.25 to 0.28 (Supplementary Table S6). KLKE and LB exhibited slightly higher H_o relative to H_e , whereas BECK showed a modest deficit. Mean Hardy-Weinberg equilibrium (HWE) P values across breeds ranged from 0.35 to 0.47. ROH and F_{ROH} metrics consistently indicated the highest autozygosity in CLH and KLKE, followed by DL,



whereas ALTs, BECK, and several other breeds showed the lowest levels (Figure 3C, D).

To complement the nuclear genome-based population structure analyses, we further investigated the maternal and paternal lineages of Xinjiang sheep using mitochondrial DNA and Y-chromosome markers. For the maternal lineage analysis, based on previously defined sheep mitochondrial haplogroups, most Xinjiang sheep individuals were assigned to haplogroups A and B, indicating that these two haplogroups represent the predominant maternal lineages in Xinjiang sheep. In addition, four individuals were assigned to haplogroup C, and one individual carried the rare haplogroup D (Supplementary Figure S4). For the paternal lineage analysis, we used a high-quality sheep Y-chromosome genome sequence as the reference and selected Y-chromosome-specific diagnostic variants for paternal haplotype assignment. Two paternal haplotypes, HY1b and HY2, were identified in Xinjiang sheep, with HY1b being the predominant paternal haplotype. Together, these results indicate that Xinjiang sheep possess diverse maternal lineages dominated by haplogroups A and B, whereas their paternal lineages are mainly represented by the common domestic sheep haplotype HY1b.

Demographic history of local sheep populations in Xinjiang

To investigate the demographic basis of population genetic structure, we used MSMC2 to infer historical effective population size (N_e) trajectories for Xinjiang indigenous sheep breeds. The inferred N_e trajectories were broadly similar among breeds until approximately 3 000 years ago (Figure 4A, C, E). Beginning at approximately 100 000 years ago, the populations underwent a pronounced bottleneck, with N_e declining to



approximately 3 000 years ago. Subsequently, breed-specific differences in N_e became evident. Estimates of population divergence indicated that breed separations in Xinjiang occurred approximately 1 000–3 000 years ago. Notably, divergence among Xinjiang breeds appeared to proceed in two phases: an earlier split at approximately 2 000 years ago involving CLH and KLKE, followed by a later split at approximately 1 000 years ago involving BSB, BYK, HSK, KEKZ, LB, TLFH, and YC (Figure 4B, D, F). We further inferred N_e trajectories using SMC++ with the larger sample set. N_e estimates from SMC++ recapitulated the overall trends observed with MSMC2 (Supplementary Figure S5). For most breeds, N_e reached a minimum at approximately 3 000 years ago. After this point, N_e trajectories diverged among breeds.

Identification of key genes associated with year-round Estrus

Xinjiang indigenous sheep exhibit marked variation in reproductive seasonality, including both year-round and seasonally estrous breeding types. To obtain a genome-wide view of genetic differentiation associated with estrus-pattern variation, we contrasted breeds exhibiting year-round estrus (CLH, HT, and DL) with breeds showing seasonal estrus (BSB, HSK, and ALTs). Across shared outlier windows between the two groups, functional enrichment analyses highlighted reproduction and endocrine-related processes (e.g., ovarian follicle development, hormonal responses, and mating behavior) and pathways linked to thyroid hormone signaling, estrogen signaling, neuroactive ligand-receptor interaction, cAMP signaling, and energy-sensing modules (PI3K-Akt, AMPK, and mTOR), collectively consistent with endocrine and energy-metabolic regulation of reproductive timing (Supplementary Tables S7–S12). Guided



by this genome-wide differentiation, we identified a highly concordant selection signature centered on *TSHR* and its flanking region on chromosome 7 ($F_{ST}=0.43$, $\theta\pi$ ratio=-1.51, XP-EHH=-9.28) (Figure 5A), nominating *TSHR* as a primary candidate locus associated with estrus-pattern divergence. This prioritization is biologically well justified, as accumulating evidence indicates that the photoperiodic neuroendocrine pathway governing seasonal breeding is predominantly mediated by the pars tuberalis-hypothalamus axis, in which *TSHR* signaling functions upstream of local thyroid-hormone activation via *DIO2*, thereby transducing day-length cues into reproductive activation or suppression (Zhou et al., 2026). Consistent with a role in distinguishing breeding types, haplotype heatmaps showed marked differentiation between year-round and seasonal-estrus groups (Figure 5B), and the focal variant at chr7:90612698 displayed striking allele-frequency contrast: the reference allele was enriched in the seasonal-estrus group, whereas the derived allele predominated in the year-round estrus group (Figure 5C). To further connect the candidate locus to its primary physiological context, we integrated multi-tissue sheep transcriptomic resources and observed that *TSHR* shows its highest expression in the thyroid gland (Figure 5D), consistent with its canonical endocrine role, and supporting tissue-relevant interpretation of regulatory variation at this locus (Gong et al., 2025). Having established *TSHR* as the top locus by both statistical and biological prioritization, we next evaluated whether chr7:90612698 represents a candidate regulatory variant. In the thyroid tissue where *TSHR* expression peaks we detected the strongest cis-eQTL signal for *TSHR* at this region, with additional nearby loci showing correlated association patterns (Figure 5E), a pattern consistent



with a regulatory haplotype affecting gene expression. Notably, in the sheep multi-tissue regulatory dataset, the derived allele at chr7:90612698 was associated with reduced *TSHR* expression in the relevant tissue (Figure 5F), providing evidence for a putative regulatory association between genotype and transcript abundance that is consistent with the observed differentiation between estrus types (Gong et al., 2025). Finally, trajectory inference suggested that the derived *TSHR* allele (g.90612698 T>G) in CLH increased rapidly approximately 2 050 years ago (Supplementary Figure S6A).

A selective sweep identifies a candidate locus for coat color variation

To investigate the genetic basis of coat color in sheep, we compared selection signals between dark-coated breeds (CLH, KLKE, TSG, and TLFH) and light-coated breeds (BYK, BECK, HT, YC, and LB). We detected strong selection signals across multiple genomic regions (Figure 6A), with the *KIT* locus on chromosome 6 showing pronounced peaks across all three statistics ($F_{ST}=0.55$, $\theta\pi$ ratio=-0.29, XP-EHH=4.42). In total, we identified 424 shared sliding-window regions and annotated 610 candidate genes associated with coat-colour variation (Supplementary Tables S13–S18). GO enrichment analysis indicated that these signals were primarily associated with melanocyte migration, differentiation, and melanin biosynthesis (Figure 6B). Haplotype heatmaps revealed marked differentiation across the *KIT* region between dark- and light-coated groups (Figure 6C). Notably, a missense variant in *KIT*, c.1 904 G>A (p.R636Q; chr6:70935444), reached its highest allele frequency in CLH, whereas it was absent in BECK and LB (Figure 6D). Cross-species sequence alignment showed that arginine at position 636 is highly conserved across mammals (e.g., cattle, goats,

and humans), whereas the R636Q substitution was observed in Xinjiang sheep (Figure 6E). Structural modelling placed this residue within a helical segment on the receptor surface; the substitution replaces positively charged arginine with neutral glutamine, suggesting that it may affect local protein conformation (Figure 6F, G). ProtParam predictions indicated that *KIT* c.1904 G>A does not alter protein length but slightly decreases molecular weight, lowers the theoretical isoelectric point, increases the instability index, and leaves overall hydrophilicity largely unchanged (Supplementary Table S18). Trajectory inference suggested a rapid rise in the derived *KIT* allele (g.70935444 G>A) in CLH approximately 480 years ago, with a comparable increase in KLKE approximately 360 years ago (Supplementary Figure S6B).

Genome-wide selection landscape identify candidate sites for tail differentiation

To investigate genetic differentiation associated with tail morphology in Xinjiang sheep, we compared selection signals between fat-rumped breeds (BSB, HSK, and BYK) and short fat-tailed breeds (YC, CLH, and HT). Genome-wide scans detected strong selection signals across multiple genomic regions. Notably, the *TBXT* locus on chromosome 8 showed pronounced peaks across all three statistics ($F_{ST}=0.17$, $\theta\pi$ ratio=1.22, XP-EHH=6.33), and additional signals were detected near *BMP2*, *LOC101117953*, and *VEGFA* (Figure 7A). Genotype-frequency analyses indicated that the AA genotype at *TBXT* g.88882077 C>A was nearly fixed in BSB, HSK, and BYK, whereas the CC genotype predominated in YC, CLH, and HT (Figure 7C). Haplotype analyses suggested that these six Xinjiang sheep breeds are dominated by two major ancestral haplotypes (Figure 7B, D). Cross-species protein sequence alignment showed



that residue 112 of *TBXT* is highly conserved as glycine (G) across mammals (goat, cattle, horse, dog, chimpanzee, human, and mouse). In contrast, Xinjiang fat-rumped sheep carry a tryptophan (W) at this site (p.G112W) (Figure 7E). ProtParam predictions indicated that *TBXT* g.88882077 C>A does not alter protein length or the theoretical isoelectric point. This substitution was predicted to slightly increase molecular weight, modestly decrease the instability index, and reduce overall hydrophobicity, with only minor effects predicted on the overall physicochemical properties (Figure 7F, G; Supplementary Table S18). In total, we identified 502 shared sliding-window regions and annotated 671 candidate genes associated with tail phenotypes (Supplementary Tables S19–S24). GO and KEGG analyses indicated enrichment of terms and pathways related to adipose development and lipid metabolism (Supplementary Tables S19–S24). In addition, trajectory inference suggested that the derived allele at *TBXT* g.88882077 C>A rose rapidly in DL and BSB approximately 1 800 years ago (Supplementary Figure S6C), indicating a recent increase in the frequency of this candidate allele rather than a direct estimate of the origin time of the fat-rumped phenotype.

DISCUSSION

Elucidating the genetic basis of phenotypic variation remains a central objective in modern genetics (Christians & Keightley, 2002; Kitano et al., 2022). Mapping genomic regions associated with economically important traits facilitates the prioritization of candidate genes and potential biological mechanisms in livestock, and provides insight into genomic regions associated with economically important traits and offers candidate



523 markers for further investigation (Saravanan et al., 2020). Here, building on previous
524 studies of Xinjiang sheep based on environmental association, SNP-array, and whole-
525 genome data (Zhang et al., 2023; Li et al., 2024), we leveraged genome-wide variant
526 maps from 1 363 individuals representing 14 indigenous Xinjiang sheep populations to
527 characterize genetic variation, population structure, genetic diversity, uniparental
528 lineages, and demographic history. The expanded sample size and breed coverage
529 allowed us to refine the population-genomic landscape of Xinjiang sheep. We further
530 integrated trait-focused selection scans with haplotype analysis, Sheep GTEx
531 expression and eQTL evidence, protein-level predictions, and Relate/CLUES-based
532 allele-frequency trajectory inference to prioritize candidate markers and genes
533 associated with coat color, reproductive traits, and tail morphotypes.

534 Nucleotide diversity predominantly ranged from 2.79×10^{-3} to 3.01×10^{-3} , consistent
535 with previous genome-wide studies (Chen et al., 2021; Li et al., 2022). For most breeds,
536 observed heterozygosity (H_o) was lower than expected heterozygosity (H_e), and overall
537 H_o was markedly lower than that reported for many overseas sheep populations
538 (Deniskova et al., 2019). CLH showed a concordant pattern of reduced π , ROH
539 accumulation, and elevated LD, consistent with a history of increased inbreeding and
540 intensified artificial selection. Because Xinjiang sheep inhabit highly heterogeneous
541 landscapes, including mountain steppe, alpine meadow, desert-oasis systems, basin
542 environments, semi-desert rangelands, piedmont areas, riparian grasslands, and plateau
543 pastures, we interpreted candidate adaptive signals in relation to breed-specific
544 ecological backgrounds rather than broad genetic or geographic groups. This revision



allows the discussion of local adaptation to better reflect the environmental heterogeneity experienced by different Xinjiang sheep breeds.

The N_e reconstructions from MSMC2 and SMC++ suggest that these demographic fluctuations temporally overlap with documented periods of intensified human migration and pastoral transitions in the region, including Bronze Age to Early Iron Age pastoral expansion in Xinjiang and adjacent Tianshan regions (Jia et al., 2020; Tian et al., 2021; Kumar et al., 2022), which may have contributed to a marked contraction of local sheep populations and setting the stage for subsequent differentiation (Zhang et al., 2017). This “post-bottleneck divergence” pattern is consistent with a reduction in cross-coalescence rates among major lineages during the last approximately 1 000–3 000 years. Meanwhile, oasis corridors, river-valley passages, and seasonal transhumance may have sustained limited gene flow at local scales, resulting in variable levels of admixture and differentiation among breeds. Historically, Silk Road associated trade networks, settlement movements, and shifts in regional governance may have altered the magnitude and direction of livestock movement, thereby accelerating recent breed differentiation and shaping the present-day genetic structure (Taylor et al., 2018). However, these temporal comparisons should be interpreted cautiously, as demographic time estimates depend on assumptions regarding mutation rate, generation time, sequencing depth, and model structure, and should not be treated as exact dates corresponding to specific archaeological or historical events. It should be noted that the MSMC2 analysis was based on the two highest-coverage individuals from each breed, a strategy that was used to improve genotype reliability and ensure



comparable analyses across breeds. However, because MSMC2 generally benefits from high-coverage genomes and larger sample sizes, the limited number of individuals and the relatively low sequencing depth of some samples may reduce the resolution of very recent demographic inference. Therefore, the MSMC2 results should be interpreted mainly as broad comparative demographic trajectories among breeds rather than precise estimates of recent effective population sizes. The largely consistent patterns inferred by SMC++ using larger breed-level sample sets provide additional support for the major demographic trends observed in this study.

For reproductive seasonality, potential selective pressures may involve both environmental cues and human-mediated selection for reproductive performance. Photoperiod is a major environmental regulator of seasonal reproduction in sheep, and seasonal variation in temperature, nutritional status, forage availability, and local management systems may further influence reproductive timing (Rosa & Bryant, 2003; Shinomiya et al., 2014). At the same time, year-round estrus and high prolificacy are economically important traits in sheep production and may have been favored by long-term artificial selection in local breeding systems. Therefore, the *TSHR*-associated signal identified in this study may reflect both photoperiod-related environmental adaptation and human selection for reproductive performance (Nakao et al., 2008a; Shinomiya et al., 2014; Karlsson et al., 2016). Using this phenotype-guided framework, selection scans identified a candidate variant at chr7:90612698 T>G. This variant was associated with altered *TSHR* expression, consistent with a potential cis-regulatory effect. It provides a testable link between selection signals, molecular consequences,



and phenotypic differentiation. Cis-regulatory effects of *TSHR* have been demonstrated in birds. Specifically, *TSHR* variation modulates photoperiodic responsiveness and influences the timing of seasonal reproductive activation in birds (Nakao et al., 2008b; Yoshimura, 2013; Karlsson et al., 2016). We therefore hypothesize that *TSHR*, a key component of the photoperiod-thyroid axis, may contribute to differences in reproductive seasonality among Xinjiang indigenous sheep, potentially through regulatory effects on *TSHR* expression. Further functional validation will be required to confirm this hypothesis.

Coat color may also have been influenced by both artificial selection and environmental factors. As a visible breed characteristic and an economically relevant trait related to wool, skin, and breed identity, coat color may have been shaped by human preferences, local aesthetic or cultural selection, and breed-standard formation (Reissmann & Ludwig, 2013; Koseniuk et al., 2018). Meanwhile, pigmentation differences may also be associated with environmental factors such as solar radiation, heat load, and climatic adaptation in small ruminants (Reissmann & Ludwig, 2013; Arenas-Báez et al., 2023). Therefore, coat-color variation in Xinjiang sheep may reflect the combined effects of human-mediated selection and potential environmental adaptation rather than a single selective force. Together, these results suggest that the *KIT* region is genetically associated with coat-color variation in Xinjiang sheep and may have been influenced by selection, although the specific selective forces remain unresolved (Sejian et al., 2018; Arenas-Báez et al., 2023). Notably, the missense variant we identified in *KIT* remained at high frequency in CLH, whereas the reference allele tended to be fixed in



white-coated breeds. This allele-frequency pattern is consistent with previous reports in pigs, cattle, horses, and multiple sheep breeds, which collectively support a key role for *KIT* in the genetic architecture of white pigmentation (Haase et al., 2007; Henkel et al., 2019). Structural modelling further supports the hypothesis that the amino-acid substitution may affect local conformation or stability of the *KIT* protein. However, whether this variant alters receptor tyrosine kinase activity, downstream signaling, or melanocyte development requires further experimental validation. This change may potentially influence receptor conformation or stability, although its effects on receptor activity, downstream signaling, and melanocyte biology require further functional validation.

In tail-type stratified comparisons, we identified a prominent selection signal at the *TBXT* locus that was consistently supported by multiple statistics (e.g., F_{ST} , $\theta\pi$ ratio, XP-EHH), suggesting directional selection associated with tail-shape divergence. *TBXT* encodes the T-box transcription factor Brachyury, a central regulator of notochord formation and caudal vertebral development. Accordingly, *TBXT* transcription is most active during early embryogenesis particularly gastrulation and axial elongation-leading to peak expression at embryonic stages. This developmental profile is consistent with the multi-tissue expression patterns observed in the sheep GTEx resource. Moreover, linked variants at this locus have been robustly associated with short-tail and fat-rumped phenotypes (Han et al., 2019; Yang et al., 2022). Consistent with these reports, the *TBXT* c.334 C>A variant occurred at very high frequency in the fat-rumped group, providing population-genetic evidence that this locus is strongly associated with



tail morphology in Xinjiang indigenous sheep. From the perspective of both artificial and environmental selection, tail fat deposition may have provided important advantages in pastoral production systems. Fat-tailed sheep can deposit large amounts of fat in the tail, sometimes accounting for more than 20% of carcass weight, and this trait may have helped meet early human demand for edible animal fat, thereby being favored by artificial selection (Xu et al., 2017). At the same time, tail fat can serve as an important energy reserve that helps sheep cope with cold stress, seasonal feed shortage, drought, long-distance movement, and harsh or resource-fluctuating environments (Alves et al., 2013; Mastrangelo et al., 2019). Therefore, the elevated frequency of the *TBXT*-associated haplotype should be interpreted as evidence of genetic association with tail morphology and fat deposition, potentially shaped by both long-term pastoral selection and environmental pressures, rather than as direct evidence for a single selective force.

Although the candidate loci identified in this study were supported by multiple lines of evidence, including population differentiation, selection scans, haplotype patterns, expression profiles, and protein-level predictions, these results do not by themselves establish causality. Therefore, independent population validation, individual-level phenotype association analyses, and functional assays will be required to confirm the biological effects of the candidate variants. In addition, because the expression and eQTL analyses were based on available Sheep GTEx resources, further studies involving more tissues, developmental stages, and environmental conditions will help clarify the regulatory mechanisms underlying these traits.

Despite these limitations, our findings have potential applications in sheep breeding and conservation. The candidate loci identified for year-round estrus, coat color, and tail morphology provide useful marker resources for future molecular breeding and marker-assisted selection after further validation. For example, *TSHR*-related variants may help improve our understanding of reproductive seasonality, whereas *KIT* and *TBXT* variants may be useful for selecting coat-color and tail-type traits according to local production goals. At the same time, the observed breed-specific differences in nucleotide diversity, LD decay, and ROH indicate that breeding programs should balance trait improvement with the maintenance of genetic diversity. In particular, breeds showing relatively high autozygosity should be carefully managed to avoid excessive inbreeding. Therefore, the genomic resources generated in this study can support both genetic improvement and the conservation of indigenous sheep genetic resources in Xinjiang.

CONCLUSION

We profiled 1 363 genomes from 14 Xinjiang sheep breeds, building a high-quality variant catalogue and integrative population-genomic framework. Phylogeny, population structure, demographic inference, and selection scans revealed non-tree-like differentiation that may be related to geography, ecology, and human-mediated connectivity. Phenotype-guided comparisons highlight major loci for key traits: *KIT* (coat color), *TSHR* (breeding seasonality), and *TBXT* (tail type), alongside endocrine and energy-metabolic candidates. These insights provide genomic resources for future

studies of trait evolution, molecular breeding, and conservation management in Xinjiang indigenous sheep.

DATA AVAILABILITY

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Zhang et al., 2025) in National Genomics Data Center (Bao et al., 2025), China National Center for Bioinformation/ Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA037360) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests

AUTHORS' CONTRIBUTIONS

C.C.L., Methodology, Visualization, Writing-original draft, Data Curation. Y.X., Formal analysis, Methodology, Visualization, Data Curation. C.Y.G., Investigation, Visualization. J.C.H., Resources, Validation. H.T., Investigation. P.F.L., Investigation. R.L., Conceptualization, Writing-Review & Editing, Supervision. Y.J., Conceptualization, Writing-Review & Editing, Supervision, Project administration.



W.W.W., Conceptualization, Writing-Review & Editing, Supervision, Project administration, Funding. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the High-Performance Computing Platform of Northwest A&F University.

REFERENCES

- Alves SP, Bessa RJB, Quaresma MAG, et al. 2013. Does the fat tailed Damara ovine breed have a distinct lipid metabolism leading to a high concentration of branched chain fatty acids in tissues? *PLoS ONE*, **8**(10): e77313.
- Arenas-Báez P, Torres-Hernández G, Castillo-Hernández G, et al. 2023. Coat color in local goats: influence on environmental adaptation and productivity, and use as a selection criterion. *Biology*, **12**(7): 929.
- Bao Y, Bai X, Bu C, et al. 2025. Database resources of the national genomics data center, china national center for bioinformation in 2025. *Nucleic Acids Research*, **53**(D1): D30–D44.
- Bi J, Yiming S, Li J, et al. 2025. Effects of INSL3 and WNT2B gene polymorphisms on seasonal reproductive traits of Xinjiang Qira black sheep, Kazakh sheep and Duolang sheep. *Animal Biotechnology*, **36**(1): 2498677.
- Browning BL, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, **103**(3): 338–348.
- Bu D, Luo H, Huo P, et al. 2021. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Research*, **49**(W1): W317–W325.
- Chan A. 2025. Silk Road orientations in Xinjiang archaeology and shifting implications for Eurasian studies. *Modern Asian Studies*, **59**(1): 1–32.



- Chen S, Zhou Y, Chen Y, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884–i890.
- Chen ZH, Xu YX, Xie XL, et al. 2021. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Communications Biology*, **4**(1): 1307.
- Christians JK, Keightley PD. 2002. Genetic architecture: dissecting the genetic basis of phenotypic variation. *Current Biology*, **12**(12): R415–R416.
- Danecek P, Auton A, Abecasis G, et al. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**(15): 2156–2158.
- De Magalhães JP, Costa J. 2009. A database of vertebrate longevity records and their relation to other life-history traits. *Journal of Evolutionary Biology*, **22**(8): 1770–1774.
- Deng J, Xie XL, Wang D-F, et al. 2020. Paternal origins and migratory episodes of domestic sheep. *Current Biology*, **30**(20): 4085-4095.e6.
- Deniskova T, Dotsev A, Lushihina E, et al. 2019. Population structure and genetic diversity of sheep breeds in the Kyrgyzstan. *Frontiers in Genetics*, **10**.
- Dominguez Mantes A, Mas Montserrat D, Bustamante CD, et al. 2023. Neural ADMIXTURE for rapid genomic clustering. *Nature Computational Science*, **3**(7): 621–629.
- Dutheil JY, Gaillard S, Stukenbrock EH. 2014. MafFilter: a highly flexible and extensible multiple genome alignment files processor. *BMC Genomics*, **15**(1): 53.
- Gong M, Zhuang Z, Sun X, et al. 2025. A multi-tissue atlas of genetic regulatory effects in sheep. <http://dx.doi.org/10.1101/2025.11.03.686346>
- 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.
- Han J, Yang M, Guo T, et al. 2019. Two linked TBXT (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. *Animal Genetics*, **50**(6): 772–777.
- Hayes BJ, Lewin HA, Goddard ME. 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics*, **29**(4): 206–214.
- Henkel J, Saif R, Jagannathan V, et al. 2019. Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. *PLoS Genetics*, **15**(12): e1008536.

753 Hoffmann I. 2010. Climate change and the characterization, breeding and conservation of animal
 754 genetic resources. *Animal Genetics*, **41**(S1): 32–46.

755 Hoffmann I. 2013. Adaptation to climate change – exploring the potential of locally adapted breeds.
 756 *Animal*, **7**: 346–362.

757 Jia P, Caspari G, Betts A, et al. 2020. Seasonal movements of Bronze Age transhumant pastoralists
 758 in western Xinjiang. *PLoS ONE*, **15**(11): e0240739.

759 Jin M, Liu G, Lu J, et al. 2024. Characteristics of genetic basis copy number variation in production
 760 and adaptation traits of Chinese indigenous sheep. *Journal of Genetics and Genomics*, **51**(5):
 761 566–569.

762 Karlsson AC, Fallahshahroudi A, Johnsen H, et al. 2016. A domestication related mutation in the
 763 thyroid stimulating hormone receptor gene (TSHR) modulates photoperiodic response and
 764 reproduction in chickens. *General and Comparative Endocrinology*, **228**: 69–78.

765 Keightley PD, Jackson BC. 2018. Inferring the probability of the derived vs. the ancestral allelic
 766 state at a polymorphic site. *Genetics*, **209**(3): 897–906.

767 Kijas JW, Lenstra JA, Hayes B, et al. 2012. Genome-wide analysis of the world's sheep breeds
 768 reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, **10**(2):
 769 e1001258.

770 Kitano J, Ishikawa A, Ravinet M, et al. 2022. Genetic basis of speciation and adaptation: from loci
 771 to causative mutations. *Philosophical Transactions of the Royal Society B*, **377**(1855):
 772 20200503.

773 Kodama Y, Shumway M, Leinonen R. 2012. The sequence read archive: explosive growth of
 774 sequencing data. *Nucleic Acids Research*, **40**(D1): D54–D56.

775 Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing
 776 data. *BMC Bioinformatics*, **15**(1): 356.

777 Koseniuk A, Ropka-Molik K, Rubiś D, et al. 2018. Genetic background of coat colour in sheep.
 778 *Archives Animal Breeding*, **61**(2): 173–178.

779 Kumar V, Wang W, Zhang J, et al. 2022. Bronze and Iron Age population movements underlie
 780 Xinjiang population history. *Science*, **376**(6588): 62–69.

781 Lawson DJ, van Dorp L, Falush D. 2018. A tutorial on how not to over-interpret STRUCTURE and
 782 ADMIXTURE bar plots. *Nature Communications*, **9**(1): 3258.

783 Leroy G, Baumung R, Boettcher P, et al. 2018. Animal genetic resources diversity and ecosystem
 784 services. *Global Food Security*, **17**: 84–91.
 785 Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree
 786 display and annotation tool. *Nucleic Acids Research*, **52**(W1): W78–W82.
 787 Li H, Handsaker B, Wysoker A, et al. 2009. The Sequence Alignment/Map format and SAMtools.
 788 *Bioinformatics*, **25**(16): 2078–2079.
 789 Li R, Yang P, Li M, et al. 2021. A Hu sheep genome with the first ovine Y chromosome reveal
 790 introgression history after sheep domestication. *Science China Life Sciences*, **64**(7): 1116–1130.
 791 Li X, He S-G, Li W-R, et al. 2022. Genomic analyses of wild argali, domestic sheep, and their
 792 hybrids provide insights into chromosome evolution, phenotypic variation, and germplasm
 793 innovation. *Genome Research*, **32**(9): 1669–1684.
 794 Li X, Yang J, Shen M, et al. 2020. Whole-genome resequencing of wild and domestic sheep
 795 identifies genes associated with morphological and agronomic traits. *Nature Communications*,
 796 **11**(1): 2815.
 797 Li Y, Li X, Han Z, et al. 2024. Population structure and selective signature analysis of local sheep
 798 breeds in Xinjiang, China based on high-density SNP chip. *Scientific Reports*, **14**(1): 28133.
 799 Lv FH, Cao YH, Liu GJ, et al. 2021. Whole-genome resequencing of worldwide wild and domestic
 800 sheep elucidates genetic diversity, introgression, and agronomically important loci. *Molecular*
 801 *Biology and Evolution*, **39**(2): 2.
 802 Manichaikul A, Mychaleckyj JC, Rich SS, et al. 2010. Robust relationship inference in genome-
 803 wide association studies. *Bioinformatics*, **26**(22): 2867–2873.
 804 Mastrangelo S, Bahbahani H, Moioli B, et al. 2019. Novel and known signals of selection for fat
 805 deposition in domestic sheep breeds from Africa and Eurasia. *PLoS ONE*, **14**(6): e0209632.
 806 McKenna A, Hanna M, Banks E, et al. 2010. The genome analysis Toolkit: a MapReduce
 807 framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**(9):
 808 1297–1303.
 809 Meadows JRS, Hanotte O, Drögemüller C, et al. 2006. Globally dispersed Y chromosomal
 810 haplotypes in wild and domestic sheep. *Animal Genetics*, **37**(5): 444–453.
 811 Nakao N, Ono H, Yamamura T, et al. 2008a. Thyrotrophin in the pars tuberalis triggers
 812 photoperiodic response. *Nature*, **452**(7185): 317–322.

813 Nakao N, Ono H, Yoshimura T. 2008b. Thyroid hormones and seasonal reproductive
 814 neuroendocrine interactions. *Reproduction*, **136**(1): 1–8.

815 Passamonti MM, Somenzi E, Barbato M, et al. 2021. The quest for genes involved in adaptation to
 816 climate change in ruminant livestock. *Animals*, **11**(10): 2833.

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

819 Puechmaille SJ. 2016. The program structure does not reliably recover the correct population
 820 structure when sampling is uneven: subsampling and new estimators alleviate the problem.
 821 *Molecular Ecology Resources*, **16**(3): 608–627.

822 Purcell S, Neale B, Todd-Brown K, et al. 2007. PLINK: a tool set for whole-genome association
 823 and population-based linkage analyses. *The American Journal of Human Genetics*, **81**(3): 559–
 824 575.

825 Reissmann M, Ludwig A. 2013. Pleiotropic effects of coat colour-associated mutations in humans,
 826 mice and other mammals. *Seminars in Cell & Developmental Biology*, **24**(6–7): 576–586.

827 Rosa HJD, Bryant MJ. 2003. Seasonality of reproduction in sheep. *Small Ruminant Research*, **48**(3):
 828 155–171.

829 Sabeti PC, Varilly P, Fry B, et al. 2007. Genome-wide detection and characterization of positive
 830 selection in human populations. *Nature*, **449**(7164): 913–918.

831 Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic
 832 trees. *Molecular Biology and Evolution*, **4**(4): 406–425.

833 Saravanan KA, Panigrahi M, Kumar H, et al. 2020. Selection signatures in livestock genome: a
 834 review of concepts, approaches and applications. *Livestock Science*, **241**: 104257.

835 Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple
 836 genome sequences. *Nature Genetics*, **46**(8): 919–925.

837 Sejian V, Bhatta R, Gaughan JB, et al. 2018. Review: adaptation of animals to heat stress. *Animal*,
 838 **12**: S431–S444.

839 Shinomiya A, Shimmura T, Nishiwaki-Ohkawa T, et al. 2014. Regulation of seasonal reproduction
 840 by hypothalamic activation of thyroid hormone. *Frontiers in Endocrinology*, **5**.

841 Speidel L, Forest M, Shi S, et al. 2019. A method for genome-wide genealogy estimation for
 842 thousands of samples. *Nature Genetics*, **51**(9): 1321–1329.



843 Stern AJ, Wilton PR, Nielsen R. 2019. An approximate full-likelihood method for inferring
 844 selection and allele frequency trajectories from DNA sequence data. *PLoS Genetics*, **15**(9):
 845 e1008384.

846 Taylor W, Shnaider S, Abdykanova A, et al. 2018. Early pastoral economies along the Ancient Silk
 847 Road: biomolecular evidence from the Alay Valley, Kyrgyzstan. *PLoS ONE*, **13**(10): e0205646.

848 Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from
 849 hundreds of unphased whole genomes. *Nature Genetics*, **49**(2): 303–309.

850 Tian D, Festa M, Cong D, et al. 2021. New evidence for supplementary crop production, foddering
 851 and fuel use by Bronze Age transhumant pastoralists in the Tianshan Mountains. *Scientific*
 852 *Reports*, **11**(1): 13718.

853 Trifinopoulos J, Nguyen LT, von Haeseler A, et al. 2016. W-IQ-TREE: a fast online phylogenetic
 854 tool for maximum likelihood analysis. *Nucleic Acids Research*, **44**(W1): W232–W235.

855 Vasimuddin Md, Misra S, Li H, et al. 2019. Efficient Architecture-Aware Acceleration of BWA-
 856 MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing
 857 Symposium (IPDPS), 314–324. Rio de Janeiro, Brazil, 2019.

858 Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from
 859 high-throughput sequencing data. *Nucleic Acids Research*, **38**(16): E164–E164.

860 Wang K, Mathieson I, O’Connell J, et al. 2020. Tracking human population structure through time
 861 from whole genome sequences. *PLoS Genetics*, **16**(3): e1008552.

862 Weir BS, Cockerham CC. 1984. Estimating f-statistics for the analysis of population structure.
 863 *Evolution*, **38**(6): 1358–1370.

864 Xu SS, Ren X, Yang GL. 2017. Genome-wide association analysis identifies the genetic basis of fat
 865 deposition in the tails of sheep (*Ovis aries*). *Animal Genetics*, **48**(5): 560–569.

866 Yang G, Wang C, Su H, et al. 2022. Development and application of a high-resolution melting
 867 analysis with unlabeled probes for the screening of short-tailed sheep *TBXT* heterozygotes.
 868 *Animals*, **12**(6): 792.

869 Yang J, Li WR, Lv FH, et al. 2016. Whole-genome sequencing of native sheep provides insights
 870 into rapid adaptations to extreme environments. *Molecular Biology and Evolution*, **33**(10):
 871 2576–2592.



872 Yao J, Chen Y, Guan X, et al. 2022. Recent climate and hydrological changes in a mountain–basin
 873 system in Xinjiang, China. *Earth-Science Reviews*, **226**: 103957.

874 Yao L, Bao A, Hong W, et al. 2019. Transcriptome profiling analysis reveals key genes of different
 875 coat color in sheep skin. *PeerJ*, **7**: e8077.

876 Yoshimura T. 2013. Thyroid hormone and seasonal regulation of reproduction. *Frontiers in*
 877 *Neuroendocrinology*, **34**(3): 157–166.

878 Zhang C, Dong SS, Xu JY, et al. 2019. PopLDdecay: a fast and effective tool for linkage
 879 disequilibrium decay analysis based on variant call format files. *Bioinformatics*, **35**(10): 1786–
 880 1788.

881 Zhang C, Zhang J, Tuersuntuoheti M, et al. 2023a. Landscape genomics reveals adaptive divergence
 882 of indigenous sheep in different ecological environments of Xinjiang, China. *Science of The*
 883 *Total Environment*, **904**: 166698.

884 Zhang J, Zhang C, Tuersuntuohe M, et al. 2023b. Population structure and selective signature of
 885 sheep around Tarim Basin. *Frontiers in Ecology and Evolution*, **11**.

886 Zhang M, Peng WF, Yang GL, et al. 2014. Y chromosome haplotype diversity of domestic sheep
 887 (*Ovis aries*) in northern Eurasia. *Animal Genetics*, **45**(6): 903–907.

888 Zhang S, Chen X, Jin E, et al. 2025. The GSA Family in 2025: A broadened sharing platform for
 889 multi-omics and multimodal data. *Genomics, Proteomics & Bioinformatics*, **23**(4): qzaf072.

890 Zhang Y, Kong Z, Yang Z. 2017. Pollen-based reconstructions of Late Holocene climate on the
 891 southern slopes of the central Tianshan Mountains, Xinjiang, NWChina. *International Journal*
 892 *of Climatology*, **37**(4): 1814–1823.

893 Zhannur N, Yuan X, Min Q, et al. 2025. Whole-genome sequencing reveals genetic architecture and
 894 selection signatures of Kazakh cattle. *Zoological Research*, **46**(2): 301–311.

895 Zhao YX, Yang J, Lv FH, et al. 2017. Genomic reconstruction of the history of native sheep reveals
 896 the peopling patterns of nomads and the expansion of early pastoralism in East Asia. *Molecular*
 897 *Biology and Evolution*, **34**(9): 2380–2395.

898 Zhou Z, Cai D, Zhao C, et al. 2026. Pan-3D Genome Analysis reveals the roles of structural variation
 899 in chicken chromatin architectures, domestication and production traits. *Advanced Science*,
 900 **13**(24): e20068.

- Zhu C, Fan H, Yuan Z, et al. 2016. Genome-wide detection of CNVs in Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP arrays. *Scientific Reports*, **6**(1): 27822.
- Zhu L, Shahbaz GM, Sun H, et al. 2025. Expression dynamics and estrogen response of estrogen receptors in Duolang sheep During Puberty. *Genes*, **16**(7): 731.
- Zhu M, Yang Y, Yang H, et al. 2023. Whole-genome resequencing of the native sheep provides insights into the microevolution and identifies genes associated with reproduction traits. *BMC Genomics*, **24**(1): 392.



Figure legends

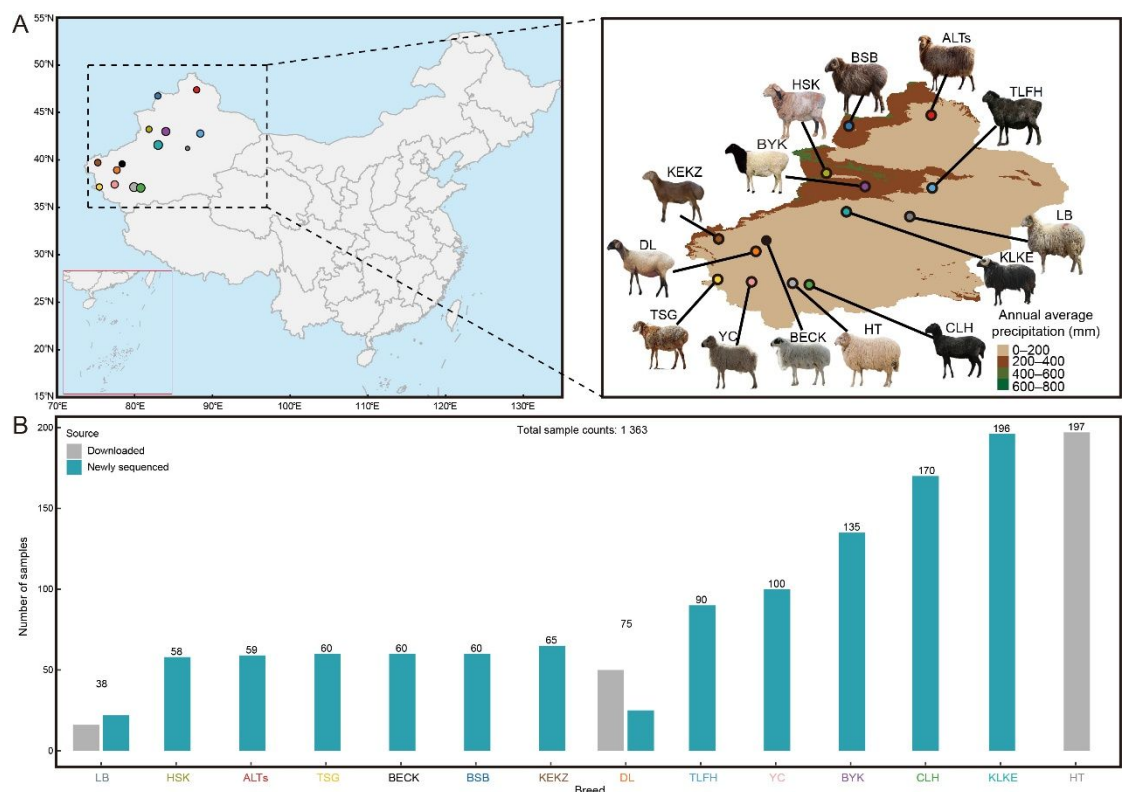


Figure 1 The geographical distribution and sample size of sheep breeds covered in this study

A: Geographic distribution of the Xinjiang indigenous sheep breeds (TLFH: Turfan Black sheep; ALTs: Altay sheep; BSB: Bashbay sheep; HSK: Kazakh sheep; BYK: Bayinbuluke sheep; KEKZ: Kirghiz sheep; DL: Duolang sheep; TSG: Tashkurgan sheep; YC: Yecheng sheep; BECK: Baerchuke sheep; HT: Hetian sheep; CLH: Cele Black sheep; KLKE: Karakul sheep; LB: Lop sheep). The background color of the sampling map indicates the predicted mean annual precipitation (mm). B: Sample composition of the 1 363 sheep genomes used in this study. Blue-green bars indicate newly sequenced samples, and gray bars indicate publicly available samples. The final dataset included 1 100 newly sequenced individuals from 13 Xinjiang indigenous sheep breeds and 263 publicly available genomes, including 197 Hetian sheep, 16 Lop sheep, and 50 Duolang sheep.

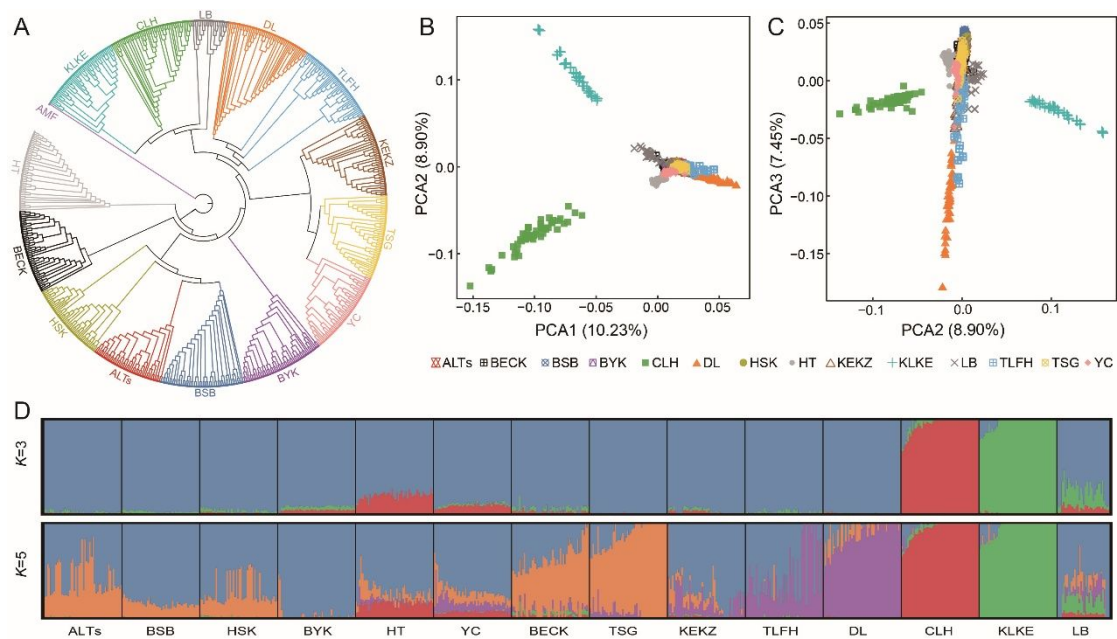


Figure 2 Population structure analysis of Xinjiang indigenous sheep

A: Neighbor-joining phylogenetic tree of sheep breeds, with Asiatic mouflon sheep as the outgroup. B, C: Principal component analysis of Xinjiang local sheep. Each color represents a breed of sheep. D: Population structure analysis with $K=3$ and $K=5$.

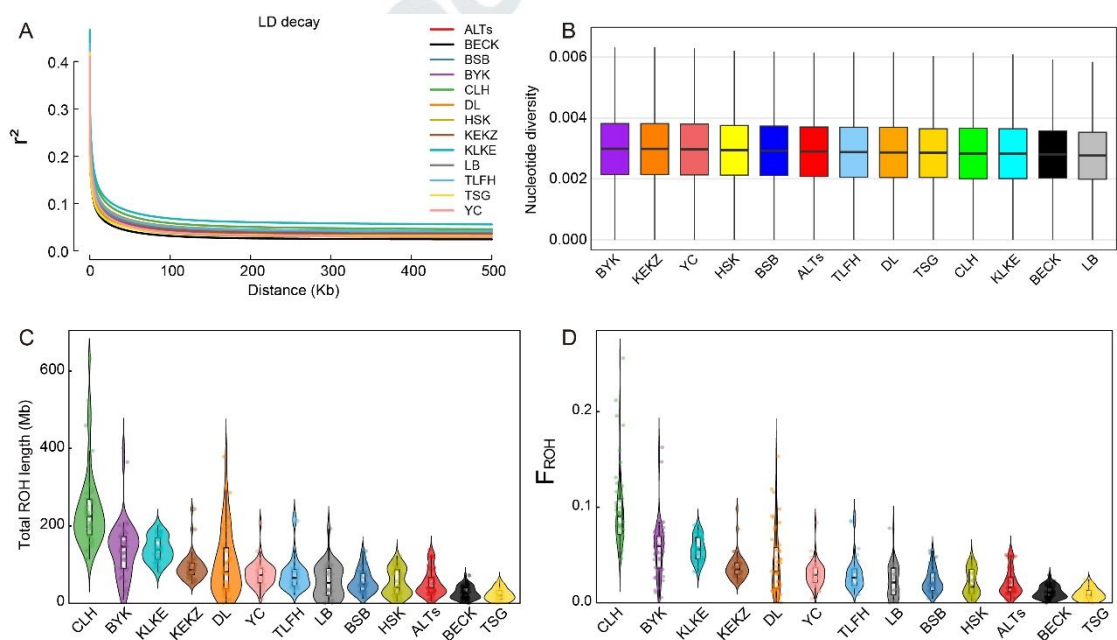


Figure 3 Diversity analysis of local sheep populations in Xinjiang

A: Linkage disequilibrium (LD) decay with genomic distance among indigenous sheep populations in Xinjiang, quantified using the squared correlation coefficient (r^2). B:

Nucleotide diversity ($\theta\pi$) measured across different sheep breed. C: Total number of purebred events for the 13 sheep breeds. D: Inbreeding coefficients (F_{ROH}) calculated for different sheep breeds.

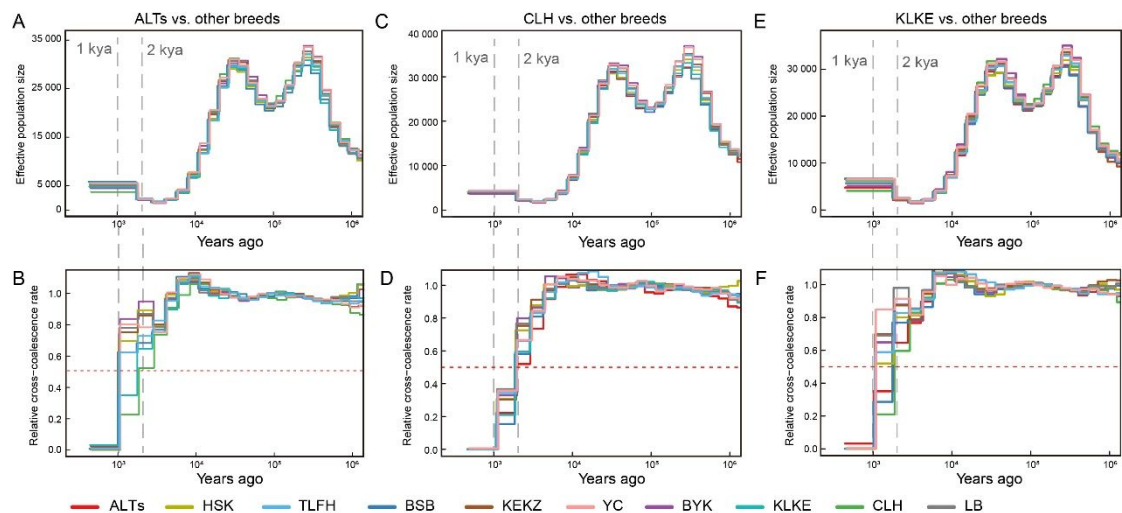


Figure 4 Demographic history of local sheep populations in Xinjiang

A: MSMC2-inferred historical effective population size (N_e) trajectories for ALTs and other breeds. B: MSMC2-inferred relative cross-coalescence rate between ALTs and other breeds. C: MSMC2-inferred historical effective population size (N_e) trajectories for CLH and other breeds. D: MSMC2-inferred relative cross-coalescence rate between CLH and other breeds. E: MSMC2-inferred historical effective population size (N_e) trajectories for KLKE and other breeds. F: MSMC2-inferred relative cross-coalescence rate between KLKE and other breeds.

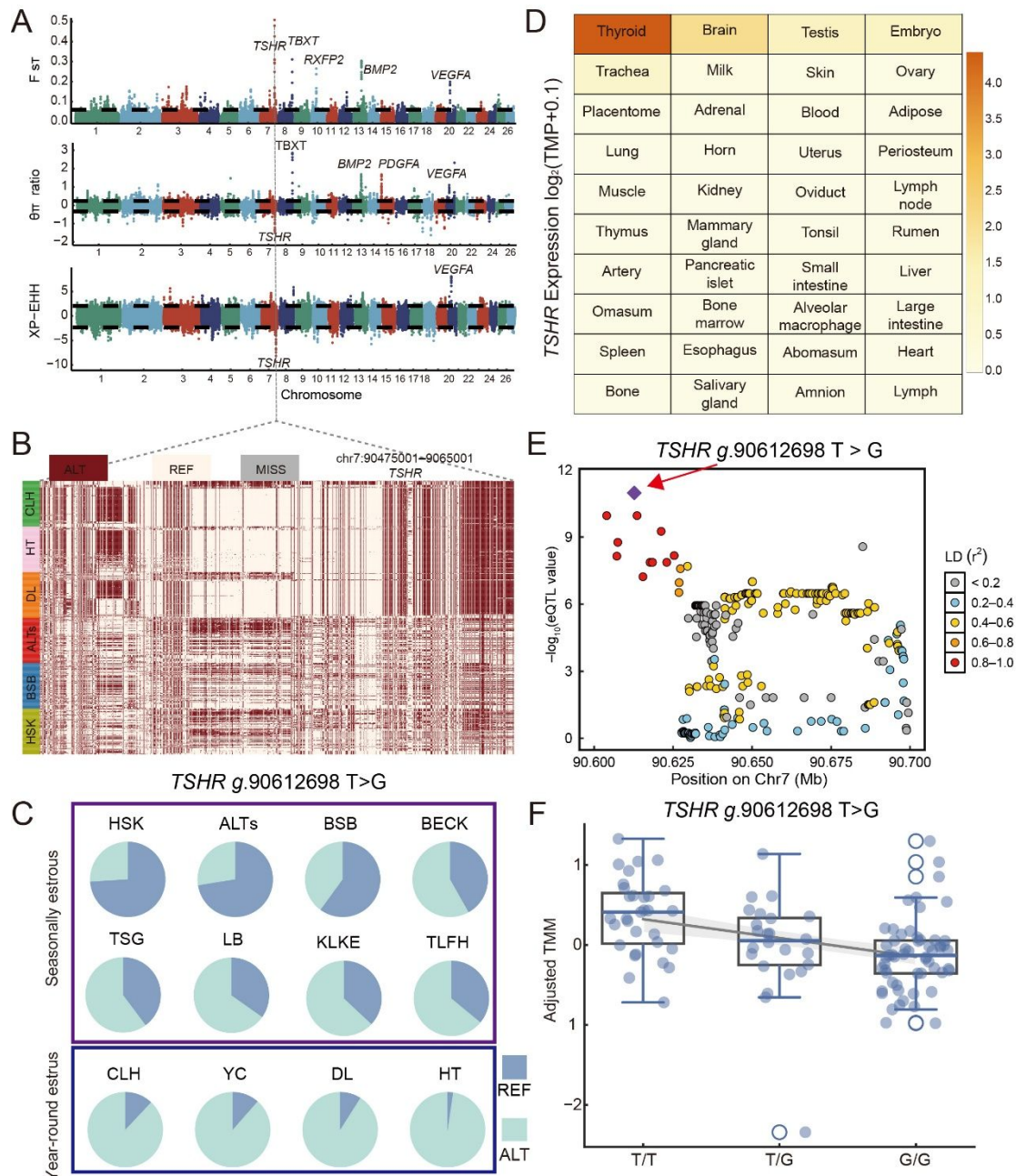


Figure 5 Selection signatures associated with reproductive seasonality

A: Genome-wide selection scans comparing year-round-estrous and seasonal-estrous sheep breeds, based on F_{ST} , $\theta\pi$ ratio, and XP-EHH methods. Overlapping windows ranked in the top 1% for at least two statistics are labeled. B: Haplotype patterns based on SNP constructs with minor allele in the chr7:90475001-9065001 bp region. C: Pie chart showing the allele frequency distribution of *TSHR* g.90612698 T>G in Twelve sheep breeds. The type of the variant allele is shown in blue and the type of the reference allele is shown in cyan. D: Expression pattern of *TSHR* gene in sheep GETx. E: Co-

localization pattern of *TSHR* g.90612698 T>G loci in sheep GETx. Target loci are indicated in purple, other different levels of interlocking are indicated by different colors. F: Expression patterns of different genotype frequencies of the *TSHR* gene.

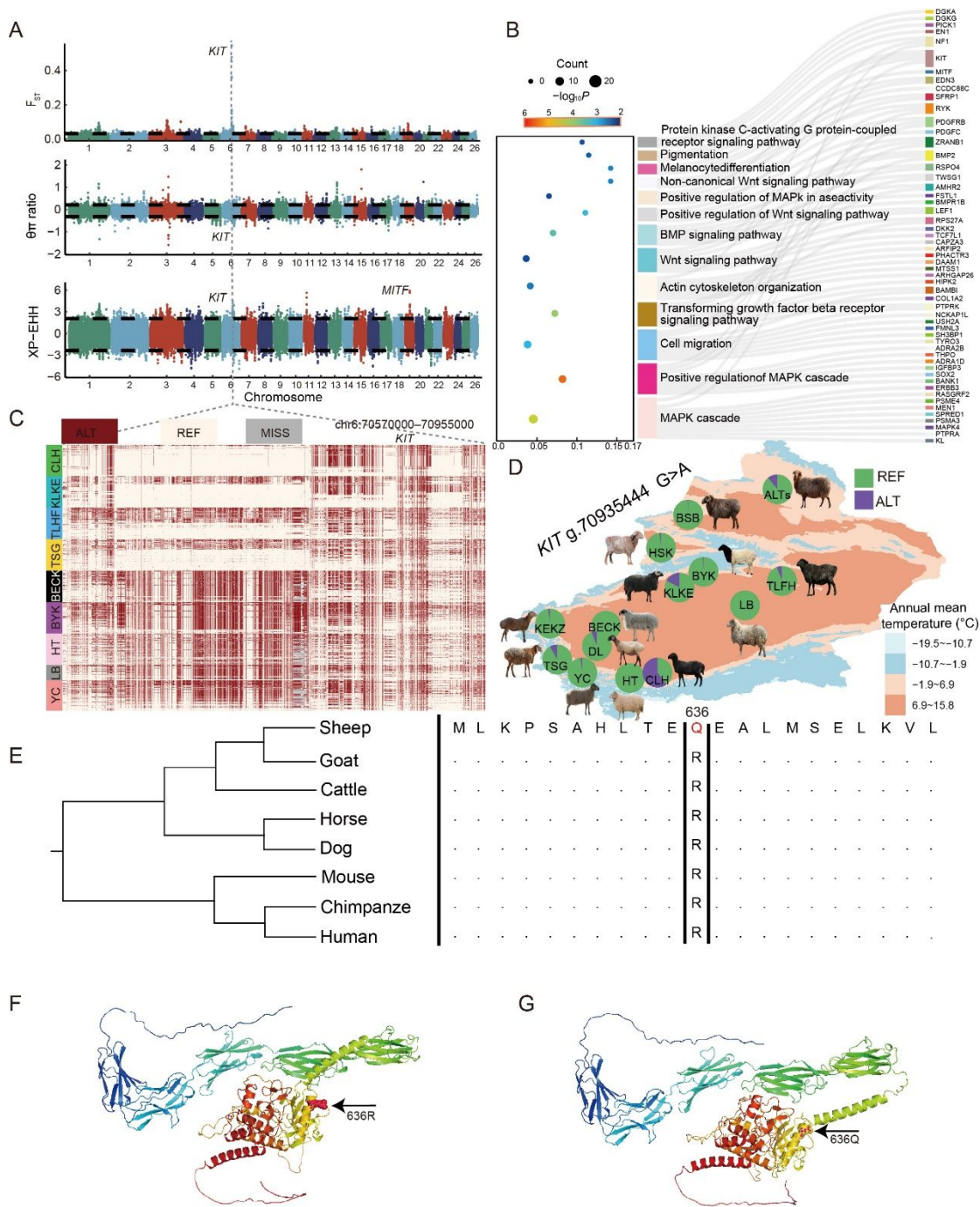
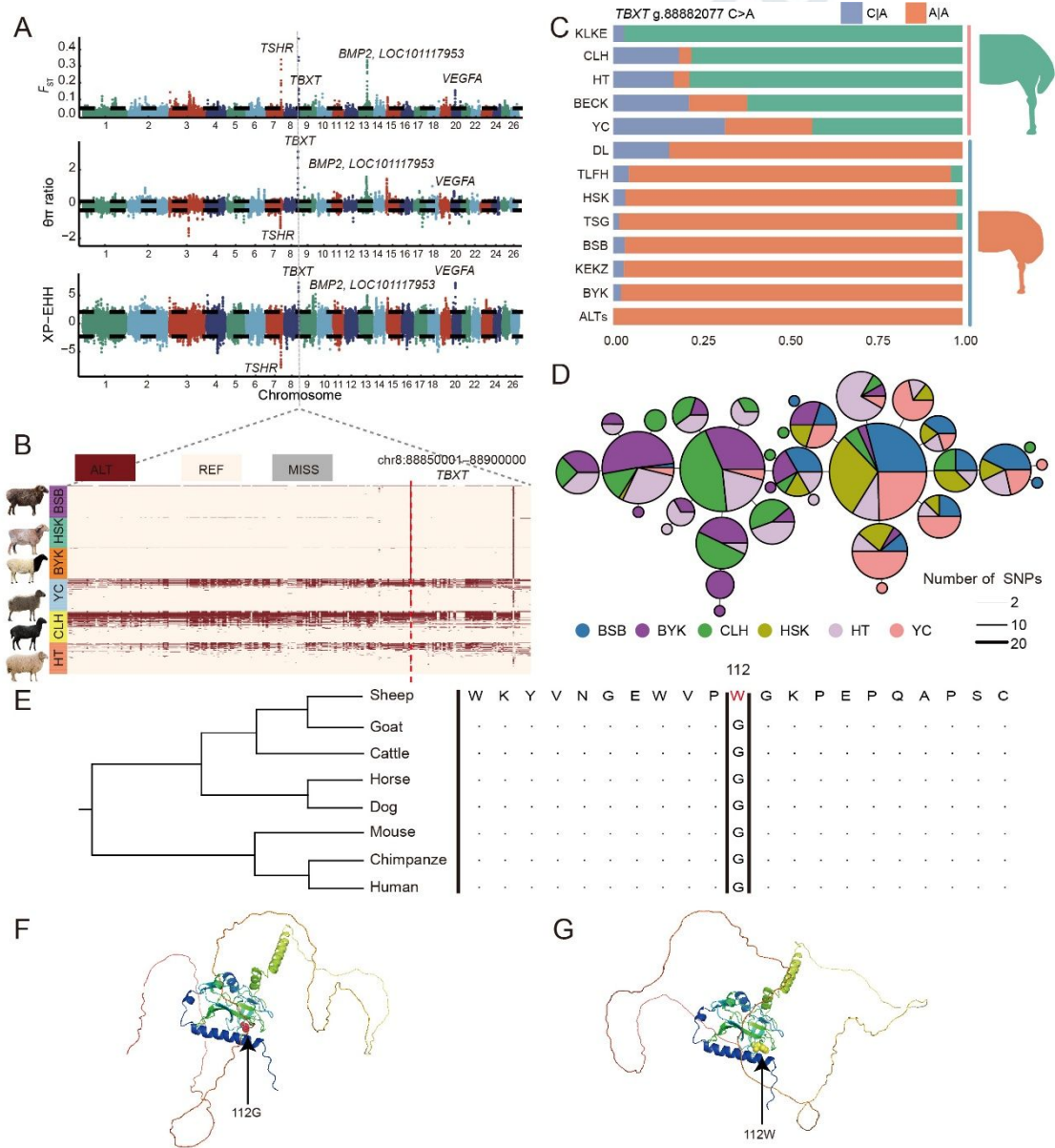


Figure 6 Selective sweeps with coat color

A: SNP-based selective sweeps with coat color, analyzed by F_{ST} , $\theta\pi$ ratio, and XP-EHH methods, with the intersections of the top 1% windows labeled. B: GO pathway

964 enrichment analyses of overlapping genes with coat color, with pathways considered
 965 significant at corrected $P < 0.05$. C: Haplotype patterns constructed from SNPs with
 966 minor allele frequencies < 0.01 in the chr 6:70570000–7095000 bp region. D:
 967 Distribution of the gene missense mutation *KIT* g.70935444 G>A across Eight sheep
 968 populations. E: Alignment of *KIT* protein sequences from nine vertebrate species. The
 969 background color of the sampling plot indicates the predicted mean annual temperature
 970 ($^{\circ}\text{C}$). F, G: Three-dimensional structures before and after mutation. (R636Q
 971 corresponds to *KIT* g.70935444 G>A). The location of mutant and natural amino acids
 972 is indicated by black arrows.



973

Figure 7 Selective sweeps with a tail type

A: SNP-based selective sweeps with a tail type, analyzed by F_{ST} , $\theta\pi$ ratio, and XP-EHH methods, with the intersections of the top 1% windows labeled. B: Haplotype patterns based on SNP constructs with minor allele in the chr 8:88850001–88900000 bp region. C: Genotype frequency distribution of the missense mutation g.88882077 C>A in the *TBXT* gene across sheep populations. D: Haplotype network based on pairwise differences within *TBXT* gene. Each circle represents a haplotype, and the size of the circle indicates the number of samples contained in that haplotype. Different colors indicate different types of sheep, where blue indicates BSB, purple indicates BYK, green indicates CLH, yellow-ish indicates HSK, indianred2-ish indicates YC, and pink indicates HT. E: Alignment of *TBXT* protein sequences from nine vertebrate species. F: and G: Three-dimensional structures before and after mutation. (G112W corresponds to *TBXT* g.88882077 C>A). The location of mutant and natural amino acids is indicated by black arrows.



990 林长春^{1, #}, 徐媛^{2, #}, 高彩月², 黄俊成¹, 哈尼克孜·吐拉甫¹, 李鹏飞¹, 李冉^{2, *}, 姜雨^{2, *}, 吴伟伟^{1, *}

991 ¹ 新疆维吾尔自治区畜牧科学院, 新疆绒毛用羊遗传育种与繁殖实验室, 农业农村部畜禽资
992 源(羊)评价利用重点实验室, 新疆维吾尔自治区绵山羊工程中心, 新疆维吾尔自治区畜牧
993 科学院畜牧研究所, 乌鲁木齐 830000, 中国

994 ² 西北农林科技大学动物科技学院, 陕西省动物遗传育种与繁殖重点实验室, 杨凌 712100,
995 中国

996 # 林长春、徐媛对本研究贡献相同。

997 * 通讯作者:

998 吴伟伟: wuweiweigp@foxmail.com;

999 姜雨: yu.jiang@nwafu.edu.cn;

1000 李冉: ran.li@nwafu.edu.cn

1001 摘要:

1002 新疆地方绵羊在复杂多样的生态环境中表现出显著的表型多样性, 但其群体结构以及重要性
1003 状的进化历史仍缺乏系统解析。本研究对来自新疆 14 个地方绵羊群体的 1 363 只个体开展
1004 全基因组重测序及群体基因组学分析。结果显示, 新疆地方绵羊可划分为 3 个主要遗传类
1005 群。其中, 策勒黑羊和卡拉库尔羊分别形成两个独立的遗传谱系, 并于约 2 000 年前与主要
1006 群体发生分化; 其余品种的分化时间相对较晚, 约为 1 000 年前。围绕繁殖季节性、毛色和
1007 脂臀等重要性状开展选择信号分析, 鉴定出多个候选基因及变异位点。全年发情与季节性发
1008 情品种的比较分析表明, *TSHR* 基因在全年发情品种中受到强烈选择。进一步精细定位筛选
1009 出候选调控变异 g.90612698 T>G, 该变异在全年发情品种中显著富集, 表达数量性状位点
1010 (eQTL) 分析显示其与甲状腺组织中较低的 *TSHR* 表达水平相关。毛色性状比较分析在策
1011 勒黑羊和卡拉库尔羊中鉴定到高频 *KIT* 基因变异 g.70935444 G>A, 等位基因频率轨迹推断
1012 显示, 该变异的频率约在 400 年前开始快速升高。对于脂臀性状, 通过祖先重组图谱重建追
1013 溯了 *TBXT* 相关变异 g.88882077 C>A 的谱系历史, 结果表明其衍生等位基因频率约在 1 800
1014 年前开始升高。综上, 本研究从全基因组水平系统揭示了新疆地方绵羊的遗传多样性、群体
1015 结构及种群历史, 并鉴定了与重要经济性状相关的候选功能变异, 为新疆地方绵羊遗传资源
1016 保护、种质创新和分子育种提供了重要的基因组学资源和理论依据。

1017 **关键词:** 全基因组重测序; 群体结构; 遗传多样性; 种群历史; 新疆地方绵羊

