



OPEN ACCESS

EDITED BY

Shi-Yi Chen,
Sichuan Agricultural University, China

REVIEWED BY

Qiuyue Liu,
Chinese Academy of Sciences (CAS), China
Boris Lukic,
Josip Juraj Strossmayer University of Osijek,
Croatia
Gong-Xue Jia,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE

Lei Gao,
✉ w.n007@163.com
Pengcheng Wan,
✉ wanpc@hotmail.com
Hua Yang,
✉ yhxjcn@sina.com
Zongsheng Zhao,
✉ zhaozongsh@shzu.edu.cn

[†]These authors have contributed equally to this work

RECEIVED 23 February 2025

ACCEPTED 25 July 2025

PUBLISHED 17 October 2025

CITATION

Gao L, Zhang Y, Zhang B, Peng W, Liu Y, Zhang Z, Wang J, Wan P, Yang H and Zhao Z (2025) Genome-wide identification of selection signals in fat-tailed and thin-tailed sheep populations.
Front. Genet. 16:1581914.
doi: 10.3389/fgene.2025.1581914

COPYRIGHT

© 2025 Gao, Zhang, Zhang, Peng, Liu, Zhang, Wang, Wan, Yang and Zhao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-wide identification of selection signals in fat-tailed and thin-tailed sheep populations

Lei Gao^{1,2*†}, Yiyuan Zhang^{2†}, Bin Zhang^{2†}, Weifeng Peng^{3†}, Yucheng Liu², Zhenliang Zhang², Jingjing Wang², Pengcheng Wan^{2*}, Hua Yang^{2*} and Zongsheng Zhao^{1*}

¹College of Animal Science and Technology, Shihezi University, Shihezi, Xinjiang, China, ²State Key Laboratory of Sheep Genetic Improvement and Healthy Production, Xinjiang Academy of Agricultural and Reclamation Science, Shihezi, China, ³College of Life Science and Agronomy, Zhoukou Normal University, Zhoukou, China

Introduction: In the evolutionary context of sheep, the development of fat tails represents an adaptive survival mechanism in response to varying food availability. Despite food resource instability, sheep store energy by accumulating tail fat to survive periods of famine. This energy storage function remains present in domesticated sheep, serving as a key evolutionary reason for the formation of sheep tail fat.

Methods: Here, we conducted whole-genome resequencing of 555 sheep samples (30 samples were newly sequenced and 525 were retrieved from published data) globally to investigate selection signatures associated with fat-tailed traits using Fixation Index (F_{ST}), Nucleotide diversity (π), cross-population composite likelihood ratio (XP-CLR), and runs of homozygosity (ROH) methods.

Result and discussion: Our examination of selection signatures in Fat-tailed and Thin-tailed Sheep Populations identified 32 candidate genes, with 6 genes (*PDGFD*, *BMP2*, *GLIS1*, *LIPE*, *MSRB3*, and *TBX15*) implicated in fat accumulation and lipid metabolism. Notably, 8 significant Gene Ontology terms (mesenchymal cell differentiation, positive regulation of ERK1 and ERK2 cascades, hormone metabolic process, nucleocytoplasmic transport, regulation of hormone levels, response to growth factor, regulation of canonical Wnt signaling pathway, and tissue morphogenesis) may play a role in fat deposition and tail fat development. These results will provide molecular targets for low-fat sheep breeding and enhance economic returns in sheep farming.

Conclusion: This study will play a crucial role in environmental adaptation and product development, comprehensively driving the development of the sheep farming industry and enhancing economic benefits.

KEYWORDS

sheep, selection signatures, fat-tail, whole-genome resequencing, gene ontology

Introduction

Sheep are among the earliest domesticated livestock, playing crucial roles in the production of meat, wool, and milk (Tian et al., 2024; Peng et al., 2022). Sheep are one of the major livestock species in China, particularly important in arid and grassland regions. Their presence not only supports the livelihoods of countless herding families in provinces such as Inner Mongolia, Xinjiang, and Qinghai, but also plays a pivotal role in the regional ecological balance. China, with a long history of domestication and breeding, boasts a rich diversity of sheep breeds that are affected by human migration and climate (Abied et al.,

2020). Additionally, sheep breeds can be classified by the type of tail fat deposition and tail morphology, including short thin-tailed, long thin-tailed, short fat-tailed, long fat-tailed, and fat-rumped sheep (Li et al., 2020). Representative Chinese breeds include short fat-tailed sheep like Small-tailed Han sheep and Mongolian sheep (Liu et al., 2016), long fat-tailed sheep like Tan sheep, Large-tailed Han sheep, and Tong sheep, short thin-tailed sheep like Tibetan sheep and Hanzhong fine wool sheep, and predominantly long thin-tailed breeds, including most fine wool and semi-fine wool sheep (Liu et al., 2021). Fat-rumped sheep have large tails with fat deposits in the rump, as seen in breeds such as Altay sheep and Kazakh sheep (Zhu et al., 2020).

It is widely believed that fat-tailed sheep have evolved from thin-tailed sheep through long-term natural selection in extremely harsh geographical and climatic conditions (Caiye et al., 2023). The fat-tail trait enables sheep to store sufficient energy, which is crucial for survival in times of food scarcity. In most regions worldwide, fat-tailed sheep are generally more adaptable to harsh environments compared to other tail types in the same area (Kalds et al., 2022). Phenotypic analyses of growth, development, and production of sheep in China, along with factors related to geographical and climatic environments, reveal that climate types significantly influence the distribution of thin-tailed and fat-tailed sheep (Abied et al., 2020). Short fat-tailed and fat-rumped sheep are primarily found in northern regions, where the fat-tail trait, allowing for greater energy storage, has evolved through natural selection in relatively harsh environments.

In contrast, southern regions are mainly populated by short thin-tailed sheep (Qi et al., 2024). The high proportion of non-terminal branched-chain fatty acids in sheep tail fat, along with its unique lipid metabolism process, suggests that genes or regulatory pathways controlling tail fat deposition may differ from those involved in fat deposition in other body parts (Alves et al., 2013). The specific genetic mechanisms underlying tail fat deposition remain unclear, with only a partial identification of candidate genes and nucleotide fragments associated with tail phenotype. Genome-wide association studies in Iranian fat-tailed and thin-tailed sheep have identified highly homozygous single-nucleotide polymorphism (SNP) loci on chromosomes X and 5 in fat-tailed sheep, and on chromosome 7 in thin-tailed sheep, suggesting a potential association of these SNPs with sheep fat deposition and meat quality (Moradi et al., 2012). Further screening has revealed a homozygous SNP site on the intron of the androgen receptor (AR) gene on the X chromosome, showing high fat-tail deposition efficiency in Altay sheep and Hu sheep, while displaying significant polymorphism in the thin-tailed Chinese Merino and Suffolk sheep, indicating a potential correlation of this SNP site with tail fat traits (Zhang et al., 2021).

In addition, by comparing the differential expression of whole-genome mRNA and conducting corresponding KEGG analysis in the fat tissues of fat-tailed Tibetan sheep and thin-tailed Dorset sheep, it was uncovered that genes with significantly altered expression levels were mostly related to lipid metabolism (Zhang et al., 2021). Using restriction fragment length polymorphism (RFLP) analysis, researchers detected two SNP loci on the X chromosome in Altay sheep and Hu sheep, which exhibit significant differences in tail fat deposition efficiency. The distribution of SNPs at the 59,571,364 and 59,912,586 loci

between sheep with significantly different tail types suggests that these two SNPs could serve as genetic markers for breeding sheep with high or low tail fat content (Zhang et al., 2021). Through whole-genome selection signal detection based on the population differentiation index F_{ST} of SNPs in Mongolian sheep and Tibetan sheep, researchers found that the expression levels of the fat deposition-related genes *PPARG* and *PDGFD* in the fat-tailed Hulun Buir sheep (divided into long-tailed and short-tailed breeds) were significantly higher than those in the short-thin-tailed Tibetan sheep (Dong et al., 2020). Additionally, within the long-tailed and short-tailed breeds, there were expression level differences in the *PPARG* gene that were positively correlated with tail fat deposition traits (Cui et al., 2022). Accumulating evidence indicates that multiple factors, including genetic variations, hormonal regulation, and environmental conditions, interact to modulate fat tail deposition in sheep. Previous studies mainly focused on individual genes or pathways, lacking a comprehensive understanding of the genetic-environmental interactions. Our study aims to bridge these gaps by integrating multiple datasets to uncover novel regulatory mechanisms underlying fat tail deposition in sheep.

The Kazakh sheep are initially from the northern foothills of the Tianshan Mountains and the southern foothills of the Altai Mountains. The region experiences dramatic seasonal climate variations, characterized by hot summers, cold winters, unpredictable spring temperatures, and rapid temperature drops in autumn. The development of Kazakh sheep is closely tied to the selective breeding practices of local ethnic groups and the stable ecological conditions (Zhu et al., 2023). Researchers at Xinjiang Academy of Agricultural Sciences conducted a breeding program by crossing multi-fetal Suffolk sheep (♂) as the paternal line with Kazakh sheep (♀) as the maternal line (Yang L. et al., 2024). Through decades of selection, they developed the hybrid F1 generation, which underwent further breeding to produce the F2 and F3 generations. The hybrid offspring exhibited phenotypic differences, notably reduced tail fat deposition (less than one-third that of Kazakh sheep). The populations of the hybrid F1, F2, and F3 generations currently consist of approximately 200 individuals each. The study involved sequencing 10 Kazakh sheep, 10 prolific Suffolk sheep, and 10 hybrid F2 generation individuals, along with resequencing data from 525 sheep previously published. This comprehensive analysis aimed to investigate the population structure and selection signals of different sheep tail types. The results are expected to elucidate the genetic regulatory mechanisms underlying tail fat deposition in sheep and offer valuable genetic resources for the molecular breeding of tail fat traits in fat-tailed sheep breeds.

Methods

Data

Thirty sheep individuals, consisting of 10 Kazakh sheep, 10 prolific Suffolk sheep, and 10 hybrid F2 generation individuals, were sourced from the sheep farm affiliated with the Xinjiang Academy of Agricultural and Reclamation Science. The collection site of the female Kazakh sheep selected for the samples is the 181th Herd Sheep Farm, which is the birthplace of the local Kazakh sheep breed. The male prolific Suffolk sheep selected for the

samples is a new variety that the Xinjiang Academy of Agricultural Sciences independently bred. The population of Kazakh sheep and prolific Suffolk sheep has reached 5,000, with about 400 in the F2 generation of hybridization. We collected 5 mL blood samples from the jugular veins of each sheep, treating them with EDTA-Na2 before promptly storing them at -20°C for future processing. DNA extraction from the blood samples was performed using the TIANamp Blood DNA Kit (TIANGEN, China), following the manufacturer's detailed instructions precisely. Following extraction, we assessed the quality of genomic DNA using 1% agarose gel electrophoresis and precisely measured its concentration and purity with the reliable NanoDrop 2000 spectrophotometer.

Read alignment and variable annotation

Moreover, to expand the sample size and thus boost the analytical reliability for both fat-tailed and thin-tailed groups, we incorporated 525 previously published sheep resequencing datasets (Li et al., 2020), with detailed information provided in [Supplementary Table S1](#). The valid sequencing data were mapped to the sheep reference genome (Oar_v4.0) using BWA (v.0.7.12) with the parameters “mem-t 4 -k 32 -M” (Li and Durbin, 2009). We enthusiastically utilized SAMtools (version 1.2) to detect Single Nucleotide Polymorphisms (SNPs), applying parameters like “mpileup-m 2 -F 0.002 -d 1,000” (Li et al., 2009). To ensure accuracy in variant calling, we thoughtfully incorporated targeted filtering criteria to minimize errors. Variant sites with a Quality by Depth (QD) value less than 2.0, a Mapping Quality (MQ) value less than 20, and a Fisher Strand (FS) value greater than 60.0 were removed. After this filtering process, the remaining variants were annotated using ANNOVAR version 21 (Yang H. et al., 2024). We applied quality filtering to the SNP dataset using VCFtools v0.1.17 (Danecek et al., 2011) excluding SNPs that satisfied any of the following criteria: (1) a call rate of 90% or lower; (2) a minor allele frequency (MAF) of 0.05 or less; or (3) a mean maximum depth below 3 or above 30. Following this quality control process, a total of 11,852,938 SNPs were kept for subsequent analyses.

Population genetic structure and LD decay analysis

We enthusiastically used the neighbor-joining (NJ) method from the Phylogeny Inference Package (PHYLIP) to create a phylogenetic tree, utilizing high-quality, filtered single-nucleotide polymorphisms (SNPs) as outlined by Felsenstein (1989). We enthusiastically explored the population structure through cluster analysis using ADMIXTURE (version 1.3.0, Alexander and Lange, 2011). It was an engaging process, as we carefully set our parameters to range from 2 to 6, executing the command “for K in 2–6, execute admixture--cv sheep.” bed | tee log.out, done” was implemented, with a cap of 10,000 iterations. In addition, to gain further insights into the genetic relationships among the samples, Principal Component Analysis (PCA) was conducted on the 555 samples. We employed the EIGENSOFT package, specifically version 7.2.1, to perform this analysis (Price

et al., 2006). We used the PopLDdecay software (version 3.4.3) (Zhang et al., 2019) to evaluate the linkage disequilibrium (LD) decay in sheep breeds. The analysis was conducted using the software's default parameters.

Selection signals analysis

The 555 sheep were categorized into two distinct groups: the fat-tailed group (including sheep with fat rumps, fat tails, long fat tails, and short fat tails) and the thin-tailed group (including sheep with long thin tails, short thin tails, and thin tails) ([Supplementary Table S1](#)). To screen for potential genomic regions, we conducted a genome-wide analysis of F_{ST} and π ratio distributions using a window-based strategy (50 kb windows with 25 kb intervals) (Danecek et al., 2011; Peng et al., 2024a). For statistical normalization, Z-score transformation was applied to F_{ST} values, while π ratios underwent log2 conversion. Candidate selection signals were defined as overlapping windows featuring the extreme values for both metrics (He et al., 2024; Yang et al., 2025). These outlier regions were then mapped to corresponding SNPs and annotated with relevant genes. In addition to the initial screening, we conducted sweep detection using the XP-CLR algorithm (version 1.0) (Chen et al., 2010). The analysis incorporated SNPs with <10% missing data (“-max-missing 0.9”) and utilized specific parameters: 1-kb grid spacing, 200-SNP maximum window size, and reduced weight for highly correlated SNPs ($r^2 > 0.95$) through (-w1 0.005 200 2000 2 -p0 0.95). Regions displaying top 1% genome-wide XP-CLR scores were designated as strong selective sweep candidates (Lv et al., 2021; Peng et al., 2025). To further identify genomic regions most frequently associated with ROH, we calculated the percentage of SNP occurrences in ROH by counting the number of times each SNP appeared in ROH across individuals using PLINK (Purcell et al., 2007) and a length threshold of ROH >0.5 Mb. This percentage was then plotted against the SNP's chromosomal position. A SNP was considered indicative of a potential ROH hotspot if its occurrence percentage exceeded 20%. Adjacent SNPs with ROH occurrence proportions above this 20% threshold formed continuous genomic segments termed ROH islands (Peng et al., 2024b).

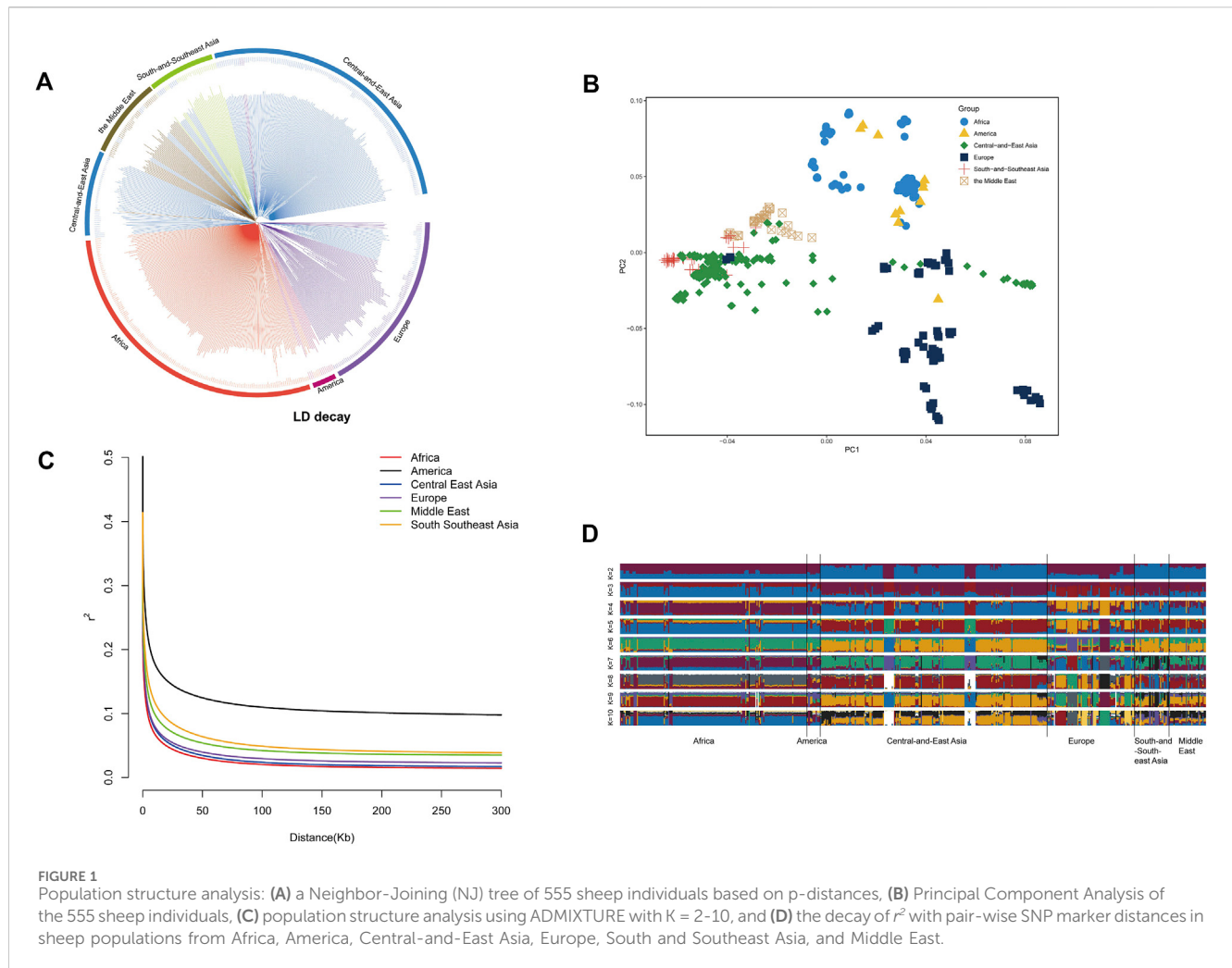
GO enrichment analysis

In addition, overlap genes identified by three methods (F_{ST} , ROH, and XP-CLR) were subjected to Gene Ontology (GO) enrichment analysis using DAVID 6.8 (Huang et al., 2009; Lukic et al., 2023). The significantly enriched GO terms ($P < 0.05$) were visualized using R software (version 4.2.1).

Results

Sequencing and mapping

Thirty Suffolk sheep were tested on the Illumina HiSeq 2500 platform, resulting in a total of 5,245.855 gigabytes (Gb) of raw data. After a stringent filtering process, 5,236.338 Gb of clean



reads were obtained. The quality of these clean reads was high, as evidenced by a Q20 value of at least 95.73% and a Q30 value of at least 89.69%. The GC content of the clean reads ranged from 43.09% to 46.62%. The clean reads were then mapped to the sheep reference genome (Oar_v4.0). The genome mapping rate varied from 98.65% to 99.32%, suggesting a strong alignment to the reference. When considering the average coverage depth across three sheep breeds, it was approximately 13.21 $\times$. Moreover, the proportion of the genome with at least 1 $\times$ average coverage exceeded 98.06%, and the proportion with at least 4 $\times$ average coverage exceeded 94.29%. These high-coverage values are indicative of the accuracy and reliability of the sequencing data. To further analyze the data, summary information was extracted from the input binary alignment/map (BAM) files using SAMTools. This tool was used to compute genotype likelihoods and convert the data into the binary variant call format (BCF). Subsequently, the ANNOVAR software was employed. It carried out the functional annotation of gene mutations and transformed the data into the Variant Call Format (VCF), which is suitable for subsequent in-depth genetic analyses. After mapping and SNP calling, a total of 11,852,938 SNPs were identified from the 555 sheep samples and can be effectively utilized to uncover genetic variations and their potential biological implications in the studied sheep breeds.

Population genetic structure and LD decay

The population genetic structure of 555 sheep was classified into 6 distinct geographic regions: the Middle East, Central and East Asia, South and Southeast Asia, Europe, America, and Africa (Figures 1A–C). The results showed that there was no apparent genetic differentiation between Asian geographical regions (including the Middle East, Central and East Asia, and South and Southeast Asia), suggesting that the sheep in these regions are influenced by human activities and there was an exchange of genetic material. In addition, sheep in the Central and East Asia region exhibited specific European lineages. For example, Chinese Merino Sheep clustered with European sheep populations. In contrast, African White Dorper sheep clustered with those in the Americas (Figure 1A). In addition, the American sheep population exhibited a high level of linkage disequilibrium (LD). This high LD level can be attributed to the intensive artificial selection and breeding strategies implemented in the American sheep industry over the past century. For instance, to meet the demand for fast growth and high meat production, breeders have consistently conducted directional selection and inbreeding mating, which has led to a reduction in genetic recombination opportunities in specific genomic regions, thereby increasing the degree of LD. The limitation of this study lies in the

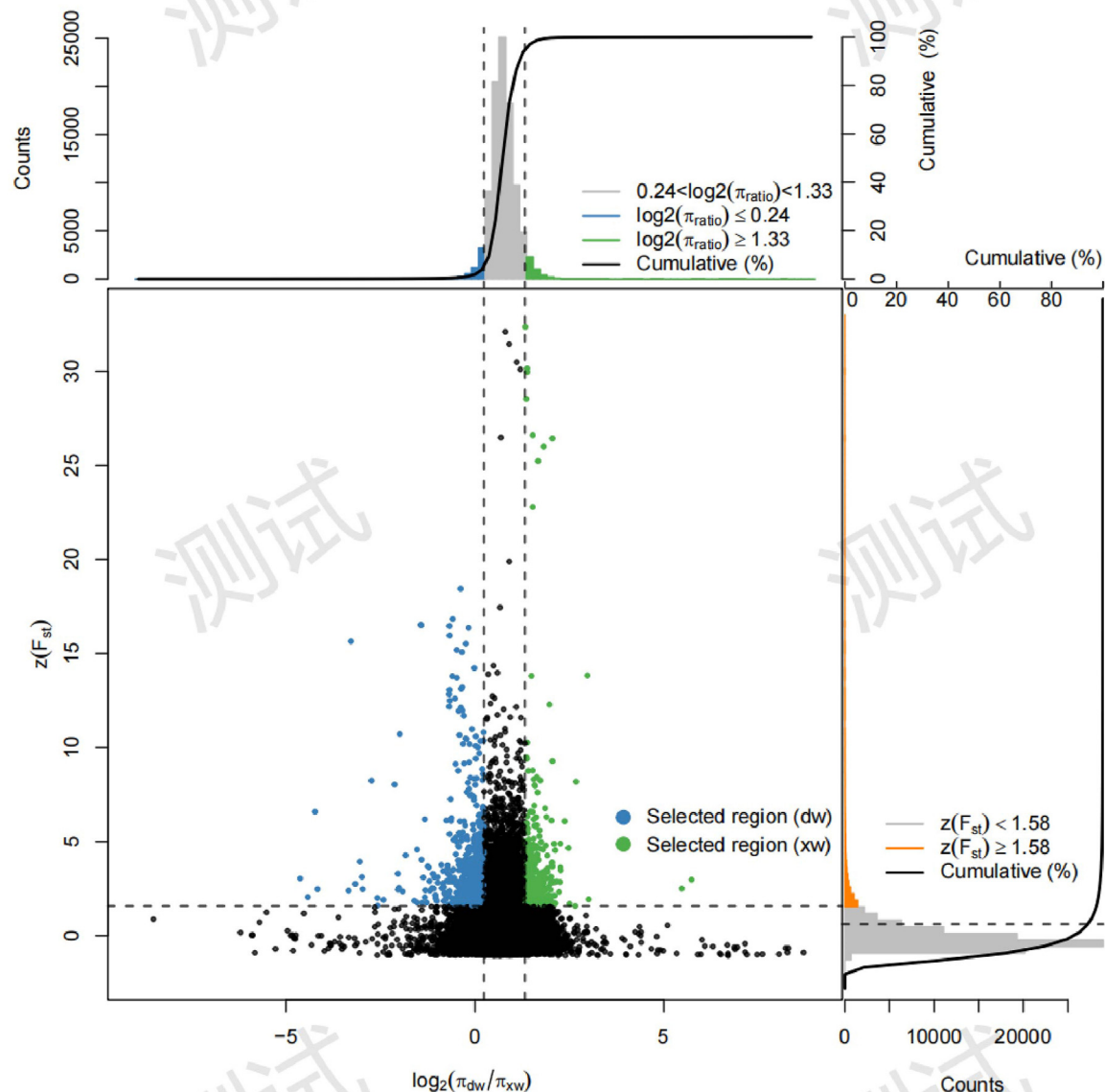


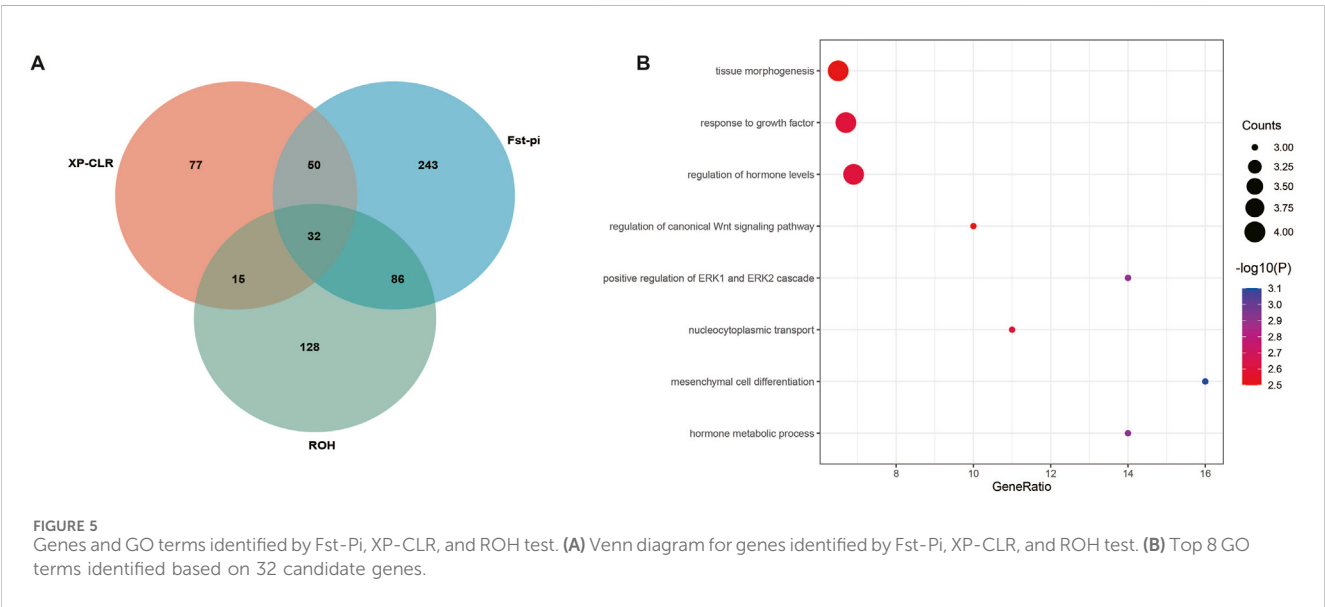
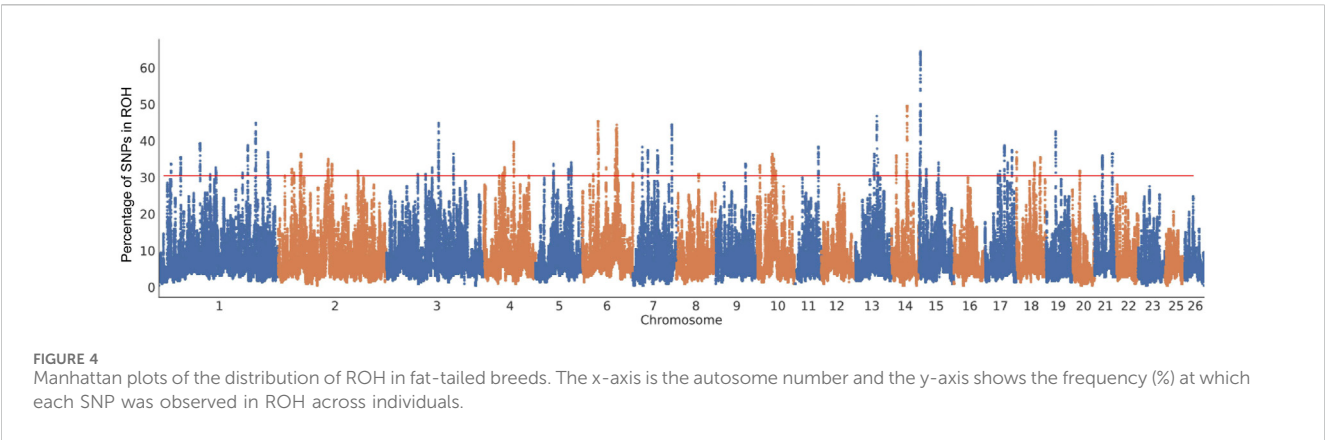
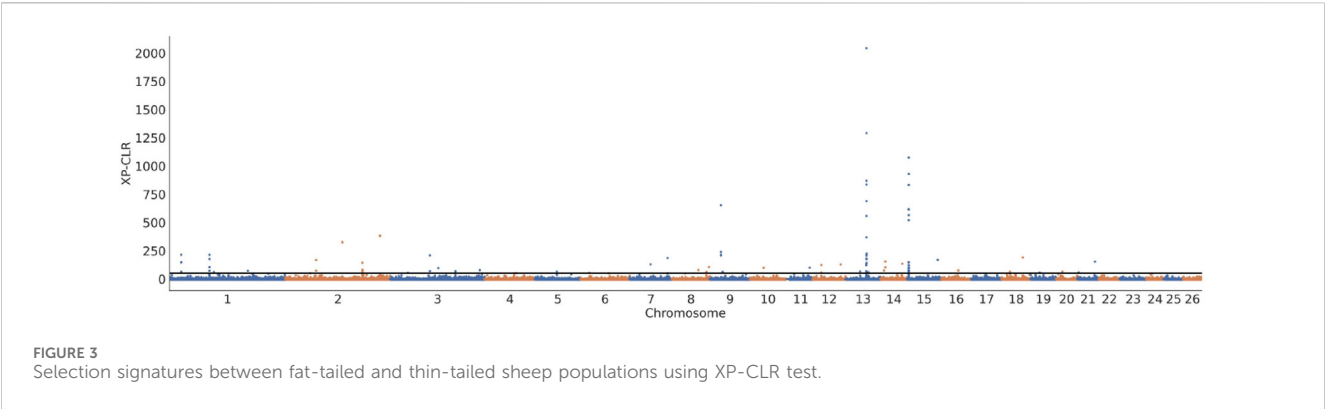
FIGURE 2
Selection signatures between fat-tailed and thin-tailed sheep populations using Fst-Pi method.

small sample size, which fails to include the original American breeds. Future research should expand the sample scope and combine genome-wide selection signal analysis to more accurately analyze the evolutionary mechanism of LD. By contrast, the African sheep population showed the lowest degree of LD (Figure 1D), indicating that it has been less influenced by artificial selection and commercial breeding practices.

Selection signals between fat-tailed and thin-tailed sheep populations

Three approaches (F_{ST} -Pi, XP-CLR, and ROH) were used to identify selection signatures between Fat-tailed and Thin-tailed

sheep populations. According to F_{ST} values with $z(F_{ST}) > 1.58$ and π ratio thresholds with $\log_2(\pi_{dw}/\pi_{xw}) < 0.24$, 411 candidate genes were identified (Figure 2; Supplementary Table S2). For the XP-CLR analysis, 169 genes were pinpointed based on the top 1% of XP-CLR values (Figure 3; Supplementary Table S3). In the ROH tests, using a threshold exceeding 30%, 256 genes were identified, with chromosome 15 harboring the highest proportion of ROH segments in animals (Figure 4; Supplementary Table S4). 32 overlapping genes were identified by the three methods (Figure 5A; Supplementary Table S5). Among these 32 genes, several have been explicitly reported in molecular mechanism studies of ovine caudal fat deposition and are implicated in adipocyte differentiation, regulation of lipid metabolism, or selection signatures for tail fat morphology. For



instance, *PDGFD* inhibits adipose tissue expansion by activating the *PDGFR β* pathway, exhibits higher expression levels in thin-tailed sheep breeds, and thereby negatively regulates caudal fat deposition. *BMP2* regulates adipocyte differentiation and lipid droplet formation, with its expression level showing a significant negative correlation with caudal fat mass. Additionally, *GLIS1* acts as a pro-

adipogenic factor, influencing mesoderm cell differentiation and caudal fat deposition. Furthermore, novel candidate genes were identified, including *TBX15*, which is involved in embryonic development regulation and known to play a role in brown adipose tissue, yet its function in ovine caudal fat remains unvalidated; and *LIPE* (Hormone-sensitive lipase), which plays a

TABLE 1 8 GO terms identified based on 32 candidate genes.

Description	-log10(P)	GeneRatio	Counts	Genes
Mesenchymal cell differentiation	3.1	16	3	BMP2 NOTCH1 ALDH1A2
Positive regulation of ERK1 and ERK2 cascade	2.9	14	3	BMP2 NOTCH1 PDGFD
Hormone metabolic process	2.9	14	3	LIPE ALDH1A2 PAPSS2
Nucleocytoplasmic transport	2.6	11	3	NOTCH1 NDC1 NUP85
Regulation of hormone levels	2.6	6.9	4	BMP2 LIPE ALDH1A2 PAPSS2
Response to growth factor	2.6	6.7	4	BMP2 MEGF8 NOTCH1 PDGFD
Regulation of canonical Wnt signaling pathway	2.5	10	3	BMP2 GNAQ NOTCH1
Tissue morphogenesis	2.5	6.5	4	BMP2 MEGF8 NOTCH1 ALDH1A2

well-established role in lipolysis, but its expression and function in ovine caudal adipose tissue have not been specifically studied or reported.

GO enrichments and KEGG analysis

Functional annotation of the 32 overlapping genes revealed eight significant Gene Ontology (GO) terms (Figure 5B; Table 1). Key terms associated with established adipogenic mechanisms include Mesenchymal cell differentiation (involving *BMP2*, *NOTCH1*, and *ALDH1A2*), which mediates precursor adipocyte maturation (Zhao et al., 2024); Response to growth factor (featuring *BMP2*, *MEGF8*, *NOTCH1*, and *PDGFD*), known to promote adipocyte differentiation through signaling cascades (Xu et al., 2023); and Tissue morphogenesis (including *BMP2*, *MEGF8*, *NOTCH1*, and *ALDH1A2*) (Jin et al., 2025), implicated in caudal fat development via vascular remodeling and adipose tissue spatial organization. Notably, novel GO terms with potential but unexplored roles in ovine fat metabolism were also identified. Positive regulation of ERK1 and ERK2 cascades modulates cell differentiation in contexts such as pulmonary fibrosis and chondrogenesis via transcription factor phosphorylation (e.g., *Runx2*); however, its functional impact on ovine adipogenesis remains uncharacterized. Similarly, the Regulation of canonical Wnt signaling—demonstrated in humans and mice to suppress preadipocyte differentiation by inhibiting *PPARγ* and *C/EBPα*—exhibits analogous regulatory potential for negative lipid deposition feedback in ovine caudal fat (Wang et al., 2025), yet, direct functional evidence remains lacking. Collectively, these findings highlight conserved adipogenic pathways while underscoring significant species-specific knowledge gaps, particularly concerning ERK/Wnt-mediated regulatory mechanisms governing ovine tail fat deposition.

Discussion

The population structure and genetic diversity of sheep play a crucial role in assessing their genetic resources, which in turn have a bearing on the utilization and conservation of these resources. After Principal Component Analysis (PCA), Neighbor-Joining tree construction, and ADMIXTURE analysis, clear differentiations

were observed among Asian (Central and East Asia, South and Southeast Asia, and the Middle East), European, and African sheep, and sheep from the American region are relatively close to Europe, and it has no obvious geographical features. Additionally, it has been noted that there is genetic interaction among sheep populations on different continents, primarily due to the impact of human activities following the domestication of sheep, particularly the use of certain commercial sheep breeds (Figures 1A–C). Especially in the Asian region, including Central-East Asia, South-Southeast Asia, and the Middle East, there have been more genetic exchanges. Because these regions are adjacent to each other, historical human activities have been frequent. For instance, during the early Neolithic Age, sheep spread from the Near East to East Asia via the Eurasian communication route, with their migration routes overlapping with historical trade routes, such as the Silk Road (Yang H. et al., 2024). The Middle East (such as the Fertile Crescent) is the origin of sheep domestication, whose genetic components gradually spread to Central Asia, South Asia, and Southeast Asia through human migration (Daly et al., 2025). Besides, we found that Chinese Merino Sheep from China were found clustered together with European sheep breeds (Figure 1A). This is because the breeding of Chinese Merino sheep is usually achieved by crossbreeding European Merino sheep (such as Spanish Merino or German Merino) with native Chinese sheep (such as Gansu mountain fine-wool sheep) (Li et al., 2023). This directed hybridization strategy directly introduced genomic fragments of European Merino sheep, resulting in a genetic structure closer to that of European breeds. LD decay revealed that African populations displayed rapid LD decay and minimal LD levels, reflecting heightened genetic recombination and reduced selection pressure. Conversely, the American population exhibited the slowest LD decay and the highest LD levels, indicating strong artificial selection and low genetic diversity (Figure 1D).

The identification of lipid selection signals in the sheep tail genome represents a significant advancement in understanding the genetic basis of fat deposition traits in sheep. In the selection of signal analysis methods, no differentiation is made between animals with long tails, short tails, or tailless characteristics. Similarly, no distinction is drawn between thin-tailed, short-thin-tailed. The classification is based on the amount of fat stored in the tail region, with individuals grouped accordingly into two categories: the fat-tailed group (encompassing sheep with fat rumps, fat tails,

long fat tails, and short fat tails) and the thin-tailed group (including those with long thin tails, short thin tails, and generalized thin tails). These signals provide crucial insights into the evolutionary and artificial selection processes that have shaped the unique fatty-tail phenotype, which is prevalent in many sheep breeds, especially in arid and semi-arid regions. Fat stored in the tail is not randomly accumulated, but rather follows a precise biological logic. In extreme temperature fluctuations, it serves as both an insulator to maintain body temperature and a metabolic reservoir that can be broken down into energy and water through β -oxidation, a dual function critical for survival in harsh climates. This trait has been sculpted by millennia of natural selection, with genomic regions governing lipid synthesis and storage becoming enriched in populations native to arid zones. Concurrently, artificial selection has amplified these signals—ancient pastoralists prioritized individuals with larger tails, as they offered better meat quality and a reliable fat source for cooking and traditional medicine, creating genetic bottlenecks that further concentrated advantageous variants. In this study, we integrated 555 genomes and conducted a comprehensive analysis of selection signals between fat-tailed and thin-tailed sheep. 32 candidate genes were identified based on three methods: Fst-Pi, XP-CLR, and ROH. Genome-wide association analysis revealed that missense mutations (G/A and C/T) of *PDGFD* (Chr15: 3900312) and *BMP2* (Chr13: 48462350) were significantly associated with tail lipid deposition (Jin et al., 2025). Functional experiments have shown that the activation and expression of *PDGFD* mutants reduce fat deposition, while *BMP2* mutants promote the differentiation of preadipocytes and increase tail fat deposition. The two genes regulate tail fat development through complementary mechanisms: *PDGFD* encourages the expansion of adipose tissue, and *BMP2* regulates energy distribution. *GNAQ* is associated with lipid metabolism and reproductive traits, and its missense mutations (such as *GPR35* g.952,651 A>G) are significantly associated with tail fat weight in Hu sheep (Zhao et al., 2024). *SPAG17* is associated with extracellular matrix (ECM) remodeling and fibrosis, and may be involved in adipocyte homeostasis regulation in caudate tissue (Xu et al., 2023). *TBX15* is related to lipid metabolism and tail shape. It was listed as a candidate gene in the genomic difference analysis of Mongolian sheep (short fat tail) and Pamei sheep (long thin tail) (Li et al., 2024). In conclusion, the formation of sheep tail lipid is regulated by a multi-gene network. The functions of core genes (such as *PDGFD*, *BMP2*, *TBX15*) have been partially clarified, but the roles of other genes (such as *GLIS1*, *MSRB3*) still need to be further explored. These findings provide a theoretical basis for molecular breeding and the improvement of tail shape (Wang et al., 2025).

Furthermore, through Gene Ontology (GO) enrichment analysis of 32 differentially expressed genes, a total of 8 significantly enriched biological processes were identified (Figure 5; Table 1). These signaling pathways can be divided into three distinct functions: (1) Mesenchymal Cell Differentiation the Cellular Foundation of Fat Accumulation. The “mesenchymal cell differentiation” process, featuring key genes *BMP2*, *NOTCH1*, and *ALDH1A2*, emerges as a central driver of tail lipid deposition. Single-cell atlas analysis reveals that mesenchymal stem cells in the tail tissues of fat-tailed sheep exhibit a biased differentiation trajectory toward adipocytes, with *BMP2* acting as a critical switch to initiate adipogenic commitment. Notably, the laminin-mediated

signaling pathway, highly expressed in Guangling large-tailed sheep and Hu sheep, reinforces this differentiation process by stabilizing the extracellular matrix (ECM) microenvironment, facilitating the transition of mesenchymal cells into mature adipocytes (Wang et al., 2025). This aligns with the observation that breeds with robust tail fat deposition show enhanced activity in mesenchymal cell lineage specification, highlighting the cellular origin of phenotypic differences; (2) Hormone Metabolic Networks systemic Regulators of Lipid Balance. Enriched GO terms related to “hormone metabolic processes” and “hormone level regulation” (encompassing *BMP2*, *MEGF8*, *NOTCH1*, and *PDGFD*) underscore the systemic regulation of tail fat deposition. These genes form a hormone-sensing network: *PDGFD* modulates insulin sensitivity in adipocytes, while *BMP2* interacts with thyroid hormone signaling to adjust lipid synthesis rates (Han et al., 2021). By fine-tuning hormone metabolism, this cluster indirectly coordinates energy allocation between tail fat storage and other physiological demands, ensuring adaptive responses to nutritional fluctuations—a mechanism particularly vital for sheep in resource-scarce arid regions; (3) Tissue Morphogenesis. Shaping the Structural Framework of Fat Depots. The enrichment of “tissue morphogenesis” terms reflects the structural remodeling required for large-scale fat accumulation. Single-cell analysis reveals that adipocytes in fat-tailed sheep secrete specific extracellular matrix (ECM) components, such as collagen subtypes, under the regulation of morphogenesis-related genes, shaping the architecture of tail adipose tissue. This structural adaptation allows for expanded lipid storage capacity, with distinct tissue morphologies (e.g., compact vs. diffuse fat distribution) corresponding to different tail shapes. The cell communication network, mediated by growth factor signals, coordinates this morphogenetic process, ensuring the synchronized expansion of fat depots and supporting the development of vasculature (Wang et al., 2025). Collectively, these findings underscore that a complex multi-gene network regulates sheep tail lipid deposition. While the functions of core genes such as *PDGFD*, *BMP2*, and *TBX15* have been partially elucidated, the roles of other candidates (e.g., *GLIS1* and *MSRB3*) remain to be characterized through functional assays, including gene knockout/overexpression studies and downstream pathway analyses. These results not only enhance our understanding of the genetic basis of fat-tailed phenotypes but also provide a theoretical framework for molecular breeding strategies that aim to tailor sheep traits to diverse agricultural and environmental needs (Wang et al., 2025).

Conclusion

This study conducted a genome-wide analysis of 555 sheep, revealing the genetic structure of the sheep population and the genetic mechanism of tail fat deposition. Three methods identified thirty-two candidate genes related to caudal lipid deposition. For instance, *PDGFD* inhibits caudal lipid deposition, *BMP2* promotes the differentiation of preadipocytes, *GLIS1* affects the differentiation of mesodermal cells, and *TBX15* and *LIPE* are newly discovered genes. GO and KEGG analyses revealed that these genes were enriched in pathways such as mesenchymal cell differentiation and growth factor response, showing that fat deposition in sheep

tails is regulated by a multi-gene network and providing a theoretical basis for molecular breeding and tail shape improvement of sheep.

Data availability statement

The data in this paper have been deposited in the Genome Variation Map (GVM) in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, under accession number GVM001174.

Ethics statement

Collection of animal samples for this study was approved by the Xinjiang Academy of Agricultural and Reclamation Sciences Ethical Committee (Number as A2024-008). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

LG: Conceptualization, Funding acquisition, Writing – review and editing. YZ: Methodology, Writing – review and editing. BZ: Data curation, Writing – review and editing. WP: Visualization, Writing – review and editing. YL: Visualization, Writing – review and editing. ZZ: Resources, Writing – review and editing. JW: Visualization, Writing – review and editing. PW: Funding acquisition, Writing – review and editing. HY: Methodology, Writing – review and editing. ZoZ: Conceptualization, Funding acquisition, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by The Science and Technology Innovation Talents Project of Corp (2023CB007-03), The National Nature Science Foundation of China (31660651), China Agriculture Research System (CARS-39-07), Young Science and Technology Top Talent Program of Tianshan Talent Training Program in Xinjiang Province (2022TSYCCX0124),

Xinjiang Agriculture Research System (XJARS-09-26), as well as Project of Corps Science and Technology in Key Areas (2024AB017).

Acknowledgments

Our heartfelt thanks go out to the researchers at our laboratories for their unwavering dedication and effort. We want to acknowledge and thank all individuals who played a part in bringing this thesis to fruition.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1581914/full#supplementary-material>

References

- Abied, A., Bagadi, A., Bordbar, F., Pu, Y., Augustino, S. M. A., Xue, X., et al. (2020). Genomic Diversity, population structure, and signature of selection in five Chinese native sheep breeds adapted to extreme environments. *Genes* 11 (5), 494. doi:10.3390/genes11050494
- Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *Bmc. Bioinform* 12, 246. doi:10.1186/1471-2105-12-246
- Alves, S. P., Bessa, R. J., Quaresma, M. A., Kilminster, T., Scanlon, T., Oldham, C., 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. doi:10.1371/journal.pone.0077313
- Caiye, Z., Song, S., Li, M., Huang, X., Luo, Y., and Fang, S. (2023). Genome-wide DNA methylation analysis reveals different methylation patterns in Chinese indigenous sheep with different type of tail. *Front. Vet. Sci.* 10, 1125262. doi:10.3389/fvets.2023.1125262
- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome. Res.* 20 (3), 393–402. doi:10.1101/gr.100545.109
- Cui, P., Wang, W., Zhang, D., Li, C., Huang, Y., Ma, Z., et al. (2022). Identification of TRAPPC9 and BAIAP2 gene polymorphisms and their association with fat deposition-related traits in Hu sheep. *Front. Vet. Sci.* 9, 928375. doi:10.3389/fvets.2022.928375
- Daly, K. G., Mullin, V. E., Hare, A. J., Halpin, Á., Mattiangeli, V., Teasdale, M. D., et al. (2025). Ancient genomics and the origin, dispersal, and development of domestic sheep. *Science* 387 (6733), 492–497. doi:10.1126/science.adn2094
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27 (15), 2156–2158. doi:10.1093/bioinformatics/btr330

- Dong, K., Yang, M., Han, J., Ma, Q., Han, J., Song, Z., et al. (2020). Genomic analysis of world wide sheep breeds reveals PDGFD as a major target of fat-tail selection in sheep. *Bmc. Genomics* 21 (1), 800. doi:10.1186/s12864-020-07210-9
- Felsenstein, J. (1989). PHYLIP-phylogeny inference package (version 3.2). *Cladistics* 5, 164–166. doi:10.1111/j.1096-0031.1989.tb00420.x
- Han, J., Guo, T., Yue, Y., Lu, Z., Liu, J., Yuan, C., et al. (2021). Quantitative proteomic analysis identified differentially expressed proteins with tail/rump fat deposition in Chinese thin- and fat-tailed lambs. *PLoS. One* 16 (2), e0246279. doi:10.1371/journal.pone.0246279
- He, S., Wang, Y., Luo, Y., Xue, M., Wu, M., Tan, H., et al. (2024). Integrated analysis strategy of genome-wide functional gene mining reveals DKK2 gene underlying meat quality in Shaziling synthesized pigs. *Bmc. Genomics* 25 (1), 30. doi:10.1186/s12864-023-09925-x
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211
- Jin, M., Liu, G., Liu, E., Wang, L., Jiang, Y., Zheng, Z., et al. (2025). Genomic insights into the population history of fat-tailed sheep and identification of two mutations that contribute to fat tail adipogenesis. *J. Adv. Res. S2090-1232* (25), 00304–2. doi:10.1016/j.jare.2025.05.011
- Kalds, P., Huang, S., Chen, Y., and Wang, X. (2022). Ovine HOXB13: expanding the gene repertoire of sheep tail patterning and implications in genetic improvement. *Commun. Biol.* 5 (1), 1196. doi:10.1038/s42003-022-04199-7
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352
- Li, X., Yang, J., Shen, M., Xie, X. L., Liu, G. J., Xu, Y. X., et al. (2020). Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat. Commun.* 11 (1), 2815. doi:10.1038/s41467-020-16485-1
- Li, C., Li, J., Wang, H., Zhang, R., An, X., Yuan, C., et al. (2023). Genomic selection for live weight in the 14th month in alpine Merino sheep combining GWAS information. *Animals* 13 (22), 3516. doi:10.3390/ani13223516
- Li, Y., Li, X., Han, Z., Yang, R., Zhou, W., Peng, Y., et al. (2024). Population structure and selective signature analysis of local sheep breeds in Xinjiang, China based on high-density SNP chip. *Sci. Rep.* 14 (1), 28133. doi:10.1038/s41598-024-76573-w
- Liu, Z., Ji, Z., Wang, G., Chao, T., Hou, L., and Wang, J. (2016). Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. *Bmc. Genomics* 17 (1), 863. doi:10.1186/s12864-016-3212-2
- Liu, J., Shi, L., Li, Y., Chen, L., Garrick, D., Wang, L., et al. (2021). Estimates of genomic inbreeding and identification of candidate regions that differ between Chinese indigenous sheep breeds. *J. Anim. Sci. Biotechnol.* 12 (1), 95. doi:10.1186/s40104-021-00608-9
- Lukic, B., Curik, I., Drzaic, I., Galic, V., Shihabi, M., Vostry, L., et al. (2023). Genomic signatures of selection, local adaptation and production type characterisation of East Adriatic sheep breeds. *J. Anim. Sci. Biotechnol.* 14 (1), 142. doi:10.1186/s40104-023-00936-y
- Lv, F. H., Cao, Y. H., Liu, G. J., Luo, L. Y., Lu, R., Liu, M. J., et al. (2021). Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression and agronomically important loci. *Mol. Biol. Evol.* 38, msab353–14. doi:10.1093/molbev/msab353
- Moradi, M. H., N Moradi ejati-Javaremi, A., Moradi-Shahrabak, M., Dodds, K. G., and McEwan, J. C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *Bmc. Genet.* 13, 10. doi:10.1186/1471-2156-13-10
- Peng, W., Zhang, Y., Gao, L., Feng, C., Yang, Y., Li, B., et al. (2022). Analysis of world-scale mitochondrial DNA reveals the origin and migration route of East Asia goats. *Front. Genet.* 13, 796979. doi:10.3389/fgene.2022.796979
- Peng, W., Zhang, Y., Gao, L., Shi, W., Liu, Z., Guo, X., et al. (2024a). Selection signatures and landscape genomics analysis to reveal climate adaptation of goat breeds. *Bmc. Genomics* 25, 420. doi:10.1186/s12864-024-10334-x
- Peng, W., Zhang, Y., Gao, L., Wang, S., Liu, M., Sun, E., et al. (2024b). Examination of homozygosity runs and selection signatures in native goat breeds of Henan, China. *Bmc. Genomics* 25, 1184. doi:10.1186/s12864-024-11098-0
- Peng, W., Zhang, Y., Gao, L., Wang, S., Liu, M., Sun, E., et al. (2025). Investigation of selection signatures of dairy goats using whole-genome sequencing data. *Bmc. Genomics* 26, 234. doi:10.1186/s12864-025-11437-9
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi:10.1038/ng1847
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795
- Qi, Y., He, X., Wang, B., Yang, C., Da, L., Liu, B., et al. (2024). Selection signature analysis reveals genes associated with tail phenotype in sheep. *Front. Genet.* 15, 1509177. doi:10.3389/fgene.2024.1509177
- Tian, Y., An, J., Zhang, X., Di, J., He, J., Yasen, A., et al. (2024). Genome-wide scan for copy number variations in Chinese Merino sheep based on ovine high-density 600K SNP arrays. *Animals* 14 (19), 2897. doi:10.3390/ani14192897
- Wang, W., Pang, Z., Zhang, S., Yang, P., Pan, Y., Qiao, L., et al. (2025). Multi-omics integrated analysis reveals the molecular mechanism of tail fat deposition differences in sheep with different tail types. *Bmc. Genomics* 26 (1), 465. doi:10.1186/s12864-025-11658-y
- Xu, Y. X., Wang, B., Jing, J. N., Ma, R., Luo, Y. H., Li, X., et al. (2023). Whole-body adipose tissue multi-omic analyses in sheep reveal molecular mechanisms underlying local adaptation to extreme environments. *Commun. Biol.* 6 (1), 159. doi:10.1038/s42003-023-04523-9
- Yang, H., Zhu, M., Wang, M., Zhou, H., Zheng, J., Qiu, L., et al. (2024a). Genome-wide comparative analysis reveals selection signatures for reproduction traits in prolific Suffolk sheep. *Front. Genet.* 15, 1404031. doi:10.3389/fgene.2024.1404031
- Yang, L., Zhang, X., Hu, Y., Zhu, P., Li, H., Peng, Z., et al. (2024b). Ancient mitochondrial genome depicts sheep maternal dispersal and migration in Eastern Asia. *J. Genet. Genomics* 51 (1), 87–95. doi:10.1016/j.jgg.2023.06.002
- Yang, C., Wang, J., Bi, L., Fang, D., Xiang, X., Khamili, A., et al. (2025). Genetic structure and selection signals for extreme environment adaptation in lop sheep of Xinjiang. *Biol. (Basel)* 14, 337. doi:10.3390/biology14040337
- Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875
- Zhang, W., Xu, M., Wang, J., Wang, S., Wang, X., Yang, J., et al. (2021). Comparative transcriptome analysis of key genes and pathways activated in response to fat deposition in two sheep breeds with distinct tail phenotype. *Front. Genet.* 12, 639030. doi:10.3389/fgene.2021.639030
- Zhao, L., Yuan, L., Li, F., Zhang, X., Tian, H., Ma, Z., et al. (2024). Whole-genome resequencing of Hu sheep identifies candidate genes associated with agronomic traits. *J. Genet. Genomics* 51 (8), 866–876. doi:10.1016/j.jgg.2024.03.015
- Zhu, C., Li, M., Qin, S., Zhao, F., and Fang, S. (2020). Detection of copy number variation and selection signatures on the X chromosome in Chinese indigenous sheep with different types of tail. *J. Anim. Sci.* 33 (9), 1378–1386. doi:10.5713/ajas.18.0661
- Zhu, M., Yang, Y., Yang, H., Zhao, Z., Zhang, H., Blair, H. T., 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. doi:10.1186/s12864-023-09479-y