Sun *et al. BMC Genomics* (2024) 25:906
https://doi.org/10.1186/s12864-024-10800-6

BMC Genomics

RESEARCH Open Access

Resequencing reveals population structure and genetic diversity in Tibetan sheep

Lixia Sun[1,2], Chao Yuan[1,2], Tingting Guo[1,2], Miaoshu Zhang[1,2], Yaqin Bai[3], Zengkui Lu[1,2*] and Jianbin Liu[1,2*]

Abstract

Background The Tibetan sheep is one of the three major primitive sheep breeds in China, representing a unique and high-quality genetic resource in the Qinghai-Tibet Plateau and neighboring high-altitude regions, exhibiting exceptional adaptability to high-altitude climatic environments. However, research on the genetic relationships among different populations of Tibetan sheep at the whole-genome level remains insufficient. This study aims to explore the population structure and historical dynamics among 11 Tibetan sheep populations, accurately assess the genetic diversity within the populations, and providing a theoretical basis for the development of targeted genetic breeding strategies for Tibetan sheep.

Results In this study, a total of 10,884,454 high-quality SNPs were obtained. All Tibetan sheep populations exhibited varying degrees of linkage disequilibrium, with similar decay rates; among them, the WT population showed the fastest decay, while the TS population exhibited the slowest decay rate. Analyses using Tajima's D and π indicated that the genetic diversity levels of the Tibetan sheep populations are generally low. Fst results revealed that most populations exhibited moderate to low levels of genetic differentiation. The effective population size among Tibetan sheep populations showed an increasing trend over time. The evolutionary relationships among Tibetan sheep populations reflect the correlation between their geographical locations and genomic genetic distances, while also indirectly confirming the impact of historical activities such as early human migration, admixture, fusion, and expansion on the population sizes and distributions of Tibetan sheep.

Conclusions The results indicate that the genetic diversity levels and genetic differentiation among Tibetan sheep populations are relatively low. In this study, we identified the genetic characteristics of Tibetan sheep populations, which exhibit low levels of diversity, genetic differentiation, and a strong population structure. A deeper genomic exploration of the population structure and diversity status of Tibetan sheep populations will provide theoretical support for subsequent genetic breeding and diversity conservation efforts.

Keywords Tibetan sheep, Population structure, Genetic diversity, Re-sequencing

*Correspondence:
Zengkui Lu
luzengkui@caas.cn
Jianbin Liu
liujianbin@caas.cn
[1]Key Laboratory of Animal Genetics and Breeding on Tibetan Plateau, Ministry of Agriculture and Rural Affairs, Lanzhou Institute of Husbandry and Pharmaceutical Sciences of Chinese Academy of Agricultural Sciences, Lanzhou 730050, China
[2]Sheep Breeding Engineering Technology Research Center of Chinese Academy of Agricultural Sciences, Lanzhou 730050, China
[3]Animal Husbandry Technology Extension Station of Gansu Provincial, Lanzhou 730050, China

Introduction

Sheep, the earliest domesticated grazing livestock, with the earliest sheep fossils discovered in the Zagros Mountains dating back approximately 11,000 years [1]. The history of sheep farming in China can be traced back to around 5,000 to 7,000 years ago based on archaeological evidence. In the early and middle Neolithic periods, wild sheep bones were predominantly found in archaeological sites, while domestic sheep bones appeared in some late Neolithic sites [2]. Based on archaeological research, a total of 145 sites containing the remains of domesticated sheep, goats, and wild sheep have been discovered in China to date. The research findings indicate an early history of sheep domestication and farming in China [3]. Chinese sheep are classified into Tibetan sheep, Mongolian sheep, and Kazakh sheep based on distribution and lineage. Tibetan sheep are one of the main domestic livestock breeds in the Qinghai-Tibet Plateau of China, serving as an important component of grassland animal husbandry in the high-altitude pastoral areas of the plateau. They are valuable genetic resources for producing high-quality wool and organic green meat under specific natural conditions and careful breeding by herdsmen [4]. Tibetan sheep not only provide high-quality protein, rich in iron, zinc, selenium, and various trace elements, which promote hemoglobin synthesis and enhance immunity, but they also offer by-products such as wool, leather, and milk, thereby increasing the income of herders [5–7].

In recent years, with the innovations in molecular biology techniques and theories, as well as the interdisciplinary integration and utilization of fields such as archaeology, anatomy, and genetics, further insights have been gained into the origin and domestication of Tibetan sheep, as well as the analysis of the population structure and diversity of Tibetan sheep [8]. In 2019, a research team successfully integrated whole-genome sequences, single nucleotide polymorphism sequences, mitochondrial DNA, and Y-chromosome sequences with archaeological information to comprehensively analyze the genetic origins and evolution of Tibetan sheep. This study revealed the evolutionary mechanism of Adaptive introgression from Argali to Tibetan sheep [9]. Furthermore, HONG et al. conducted genome resequencing analysis on 1098 domestic sheep from 154 breeds and 69 wild sheep from 7 breeds, revealing the genomic introgression map from wild sheep to domestic sheep. They discovered that the *MSRB3* gene haplotype responsible for ear width originated from wild Argali sheep and is widely present in Tibetan sheep such as the Oula breed in the Qinghai-Tibet Plateau region [10]. The identification of the *VPS13B* gene locus associated with facial contour through haplotype introgression from Argali and Iranian Mouflon sheep into the Tibetan sheep genome reveals the role of gene introgression from wild relatives in the domestication and improvement process. This provides new insights into the domestication, migration, and phenotypic diversity of sheep [10].

To date, no research team has conducted such a comprehensive analysis of genetic diversity, effective population size, and population structure of multiple Tibetan sheep populations in the Tibetan Plateau region using resequencing technology. This study aims to elucidate the mechanisms of population genetic evolution in 11 Tibetan sheep populations and, based on the research findings, develop strategies for the conservation and innovative utilization of genetic resources.

Materials and methods

Sample collection, DNA extraction and sequencing

A total of 220 blood samples were collected from 11 Tibetan sheep populations near the mountain ranges of the Tibet sheep in China (sampling locations is shown in Table S1). Each sample was collected from the jugular vein, with 5 mL of blood stored in EDTA tubes. Subsequently, DNA was extracted, and its integrity and concentration were analyzed. The genomic DNA was randomly fragmented using enzymes into short DNA fragments, followed by end repair, dA-tailing, and adapter ligation. The DNA fragments were purified using AMPure XP beads, and fragments ranging from 300 to 400 bp were amplified by PCR. The constructed libraries underwent purification, quality control, and were then subjected to sequencing on the Hiseq X10 PE150 platform, with a sequencing depth of approximately 5X.

Data quality control and genome alignment

The raw image data obtained from sequencing was converted to sequence data through base calling, and the results were stored in FASTQ file format. Strict filtering was applied to clean reads, and subsequent analysis was conducted using high-quality clean reads. Quality Trim was utilized for quality control of short sequences, filtering out reads containing adapters, reads with single-base N percentage exceeding 10% in single-end reads, and low-quality reads with more than 50% of base quality values $Q \leq 20$ in single-end reads.

The filtered high-quality reads were aligned to the genome of the Tibetan sheep, compiled by the Sheep Resources and Breeding Innovation Team of the Chinese Academy of Agricultural Sciences, using BWA software [11]. Duplicate sequences of the same short sequence from the sequencing process were removed using Picard software (http://sourceforge.net/projects/picard/). Given the high mismatch rate near InDels, which may be mistakenly identified as SNPs, the GATK software was employed to locally realign the areas around variations (typically InDels) to obtain accurate variant information [12]. Additionally, base quality recalibration was

performed using GATK to bring the quality values of reads in the final output BAM file closer to the true mismatch probability with the reference genome, thus attaining high-quality and reliable variants. Bedtools (v2.27.1) was utilized for genome coverage and sequencing depth statistics [13].

Mutation detection and annotation

Variants refer to modifications at the genomic level caused by the insertion or deletion of single or several nucleotides, resulting in DNA sequence polymorphism. Prior to variant detection, data processing is required, which generally includes local realignment and base quality recalibration.

(1) Local Realignment: Due to the presence of numerous mismatches in regions near insertions or deletions (InDels), these can often be misidentified as single nucleotide polymorphisms (SNPs). To obtain accurate variant information, we utilize GATK to perform local realignment in the vicinity of the variants (typically InDels).
(2) Base Quality Re-calibration: To reduce sequencing errors and ensure that the base quality values in the final output BAM file better reflect the true probabilities of mismatches with the reference genome, we employ GATK for base quality recalibration. This process ultimately yields high-quality and reliable variants.

We apply the UnifiedGenotyper module of GATK (version 3.4–46) on the processed alignment files to perform variant detection across multiple samples. The detected variants are filtered using VariantFiltration with the following parameters: -Window 4, -filter "QD<4.0 || FS>60.0 || MQ<40.0", and -G_filter "GQ<20" (where QD represents Variant Confidence/Quality by Depth; FS indicates the Phred-scaled p-value derived from Fisher's exact test for strand bias detection; MQ stands for RMS Mapping Quality; GQ signifies Genotype Quality). We utilize ANNOVAR for the functional annotation of the detected variants [14].

Analysis of genetic diversity

Genetic diversity and population differentiation were assessed by calculating whole-genome nucleotide diversity, Tajima's D, and genetic parameter F_{ST} [15]. The genetic differentiation between populations was evaluated using the VCF Tools software to compute pairwise F_{ST} values between different populations [16]. F_{ST} values range from 0 to 1, where a higher value indicates greater differentiation and genetic divergence: 0 implies random mating between populations with no genetic differences, while 1 signifies complete isolation and completely different genotypes between populations. Nucleotide diversity π was computed using Perl programming, based on the formula for calculating π, $\pi = \sum x_i x_j \pi_{ij}$, where X_i and X_j represent the frequencies of the i-th and j-th types of DNA sequences, and π_{ij} indicates the ratio of different nucleotides between the i-th and j-th types of DNA sequences (perl: https://github.com/xiekunwhy/gindex) [17]. Tajima's D values for each population were calculated using VCFTools (parameters setting: Slip window 100 Kb, Step 10 Kb) [16].

Linkage disequilibrium (LD) and effective population size analysis

Based on the different allele frequencies and genotype frequencies within each population, the extent of linkage disequilibrium measured by the coefficient r^2 was assessed at various distances. The average r^2 values within populations were calculated for different marker combinations at different distances, and a linkage disequilibrium decay map for 11 Tibetan sheep populations was created based on the mean r^2 values corresponding to different distances. For each population, the top 10 individuals with the deepest sequencing coverage were selected, and SMC^{++} was utilized to analyze the population history dynamics, evaluating changes in population size over time [18]. Use the estimate function of SMC^{++} software to estimate the population historical dynamics of each population, with the main parameters set to - p 0.5-m 1.51e-8 w 100-em 20 sp cubic, and other parameters using default values. The neutral mutation rate was set at $1.51*10^{-8}$, with a generation time of 3 [19–21].

Analysis of population historical dynamics

To investigate the population structure and genetic evolutionary relationships of the 11 Tibetan sheep populations, PCA analysis was conducted using the GCTA software to perform dimensionality reduction on the sample data [22]. Orthogonal transformation was applied to convert a series of potentially linearly correlated variables into a set of linearly uncorrelated new variables, known as principal components, enabling the representation of the data's characteristics in a smaller dimensional space. The results of the PCA analysis were visualized using the Ggplot2 package in the R software (https://yanglab.westlake.edu.cn/software/gcta/#PCA).

For the construction of the ML tree, SNP data was first transformed into Treemix format data (https://bitbucket.org/nygcresearch/treemix/wiki/Home), Taking goat as the outgroup (Browse - GSA - CNCB-NGDC: CRR865811). The Treemix software was then utilized to build a maximum likelihood tree of the populations with 500 rounds of Bootstrap, and Gotree was used to integrate the Bootstrap tree with the original tree to obtain support information (Bootstrap values) for nodes in the

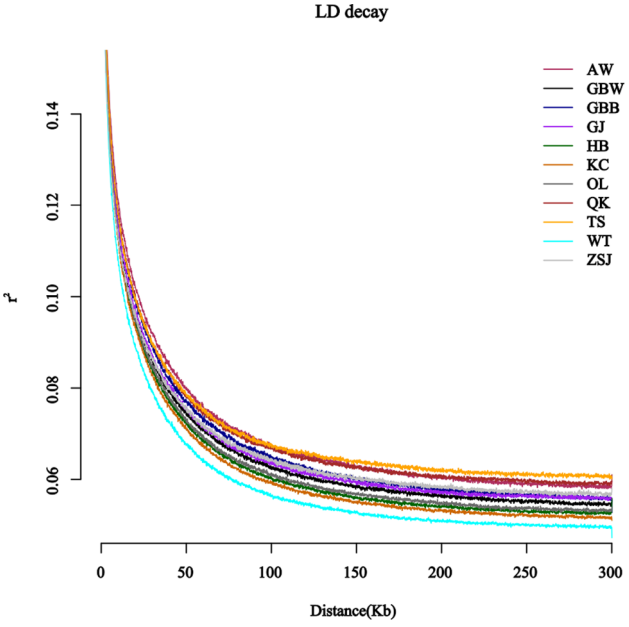

Fig. 1 LD decay map of Tibetan sheep populations

original tree (https://github.com/evolbioinfo/gotree). The Structure software was employed for population structure component analysis, assuming ancestral population numbers (subgroup numbers) ranging from K=2 to K=9. Cross-validation was performed to estimate individual ancestry and admixture proportions, and to calculate the Cross Validation Error at different K values to determine the optimal K value. The results of the population structure analysis were visualized and plotted using the R software.

Results

Sequencing data and genome-wide genetic variation

After variant detection and quality control, a total of 10,884,454 SNP loci were identified across 220 Tibetan sheep individuals in this study. Among these, HB detected the highest number of SNPs, while OL detected the fewest. The genomic variant annotation results are presented in S6, indicating that genomic variants predominantly occur in intergenic regions, followed by intronic regions. Variants are less frequent in some overlapping regions, such as exonic regions and regions within 2 bp of splice sites, 5'UTR and 3'UTR overlapping regions, as well as upstream and downstream overlapping regions.

Linkage disequilibrium and effective population size estimates

Linkage disequilibrium analysis of population genomes was conducted to investigate the relationship between average Linkage disequilibrium coefficients and marker spacing among the 11 Tibetan sheep populations at different genomic loci. The degree of linkage disequilibrium was assessed using the Linkage disequilibrium correlation coefficient (r^2) between adjacent genomic loci. The results indicated that all 11 populations exhibited some level of linkage disequilibrium, with relatively similar decay rates. Among them, WT showed the fastest decay, while TS exhibited the slowest decay rate, suggesting that WT had a smaller LD decay distance compared to other Tibetan sheep populations, indicating slightly higher genetic diversity than other populations (Fig. 1).

The SMC^{++} analysis results (Fig. 2) revealed that the effective population sizes of the 11 Tibetan sheep populations generally showed an increasing trend between generations. Around 8000 to 10,000 years ago, all Tibetan

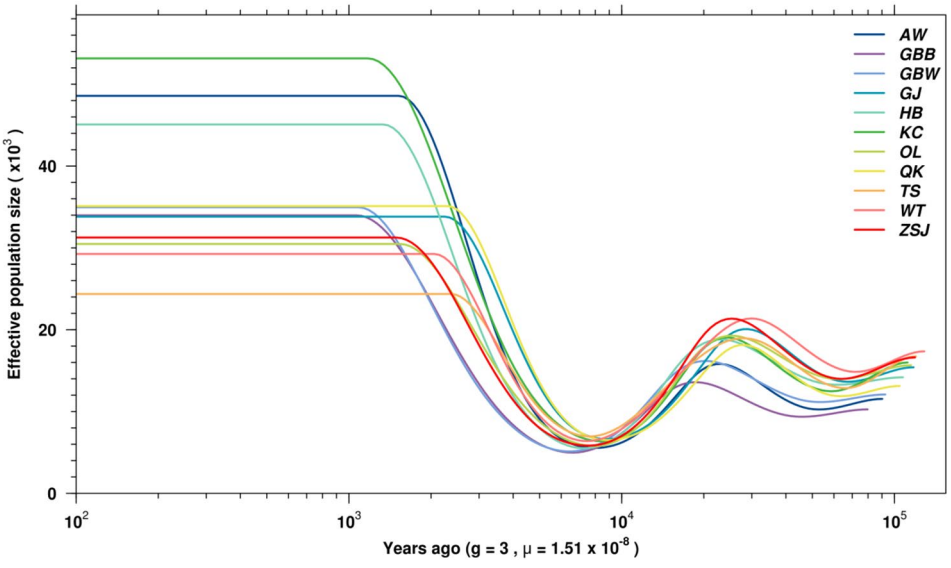

Fig. 2 SMC^{++}infers the effective population size of Tibetan sheep populations

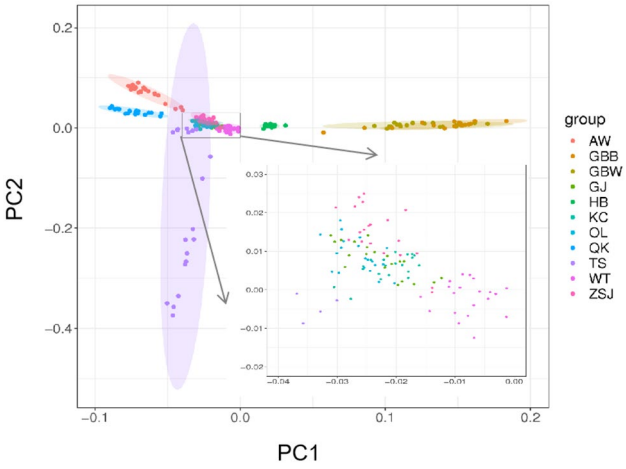

Fig. 3 Principal component analysis of Tibetan sheep populations

sheep populations experienced a reduction in effective population size, reaching historical lows, followed by a gradual increase until the effective population size stabilized. The occurrence of population bottlenecks due to factors such as domestication or historical events led to a decrease in effective population sizes, with the sizes continuing to rise following the completion of Tibetan sheep domestication. As Tibetan sheep domestication concluded and expanded gradually with agricultural development, variations in factors such as population gender ratios, genetic diversity, reproductive rates, and genetic drift resulted in differing final effective population sizes. Larger effective population sizes lead to a faster accumulation rate of new mutations, indicating greater population diversity. Among the populations, KC exhibited the largest effective population size, followed by AW and HB, while TS had the smallest effective population size. Consequently, KC possessed the highest genetic diversity, followed by AW and HB, with TS exhibiting the lowest genetic diversity.

Genetic structure and diversity of tibetan sheep population

To investigate the genetic structure and phylogenetic relationships among the 11 Tibetan sheep populations, principal component analysis (PCA) was conducted. The results revealed that the first principal component (PC1) separated HB, ZSJ, and QK, while the second principal component (PC2) distinguished AW, QK, and TS. Additionally, a closer genetic distance was observed between QC, TS, GJ, OL, and ZSJ (Fig. 3).

Using whole-genome SNPs for constructing a Maximum Likelihood (ML) tree, the evolutionary relationships among the 11 Tibetan sheep populations were investigated. The WT population belongs to the elderly group among the 11 populations studied. The AW and ZSJ populations formed a distinct branch earliest, followed by OL, QK, KC, and GJ. The GBB and GBW populations appeared as sister groups in the most recent differentiation events (Fig. 4). The relationships among the Tibetan sheep populations in Gansu, including GJ, KC, QK, and OL, are closer. Interestingly, the TS population from Gansu is more closely related to the populations of GBB, GBW, and HB from Tibet, while the ZSJ population from Qinghai has a closer relationship with the AW population from Tibet. These findings corroborate the results of the PCA analysis (Fig. 5).

To further investigate the population structure, this study conducted STRUCTURE analysis. When K=2, it is evident that the majority of the 11 Tibetan sheep populations are formed through hybridization of two ancestral

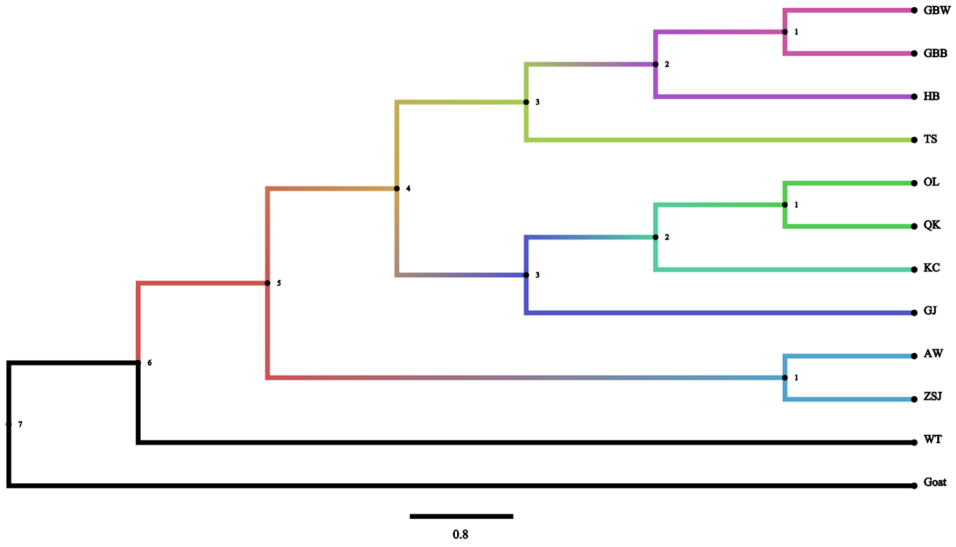

Fig. 4 Phylogenetic trees of Tibetan sheep populations

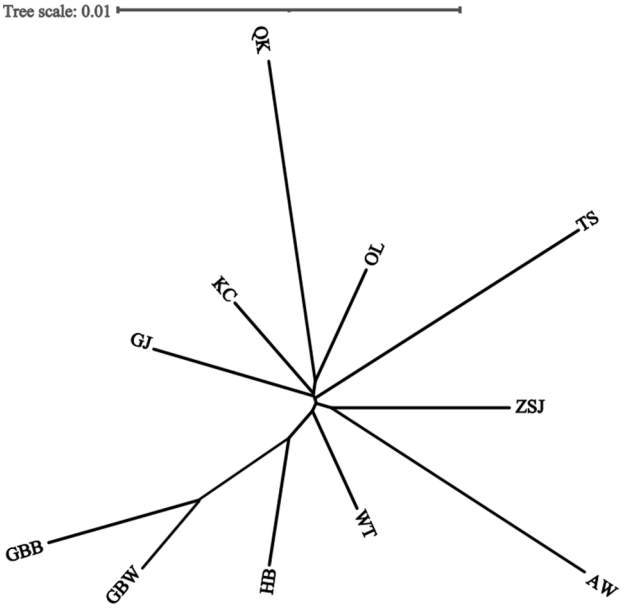

Fig. 5 Phylogenetic trees of Tibetan sheep populations (Rootless tree)

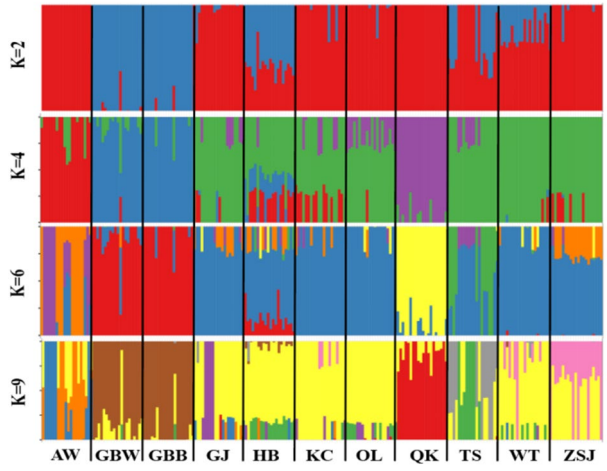

Fig. 6 Genetic structure analysis of Tibetan Sheep population

subgroups. The populations of AW and QK exhibit pure lineage without the admixture of another subgroup. When K=4, based on differences in ancestral component compositions, the 11 populations can be distinctly classified into four subgroups: GBB and GBW form one subgroup, while GJ, HB, KC, OL, TS, WT, and ZSJ form a separate subgroup, and AW and QK are each in their own independent subgroups. From K=6 to K=9, the ancestral lineage compositions of the 11 Tibetan sheep populations gradually vary, leading to population differentiation among the different populations, albeit with varying degrees of genetic admixture (Fig. 6).

To investigate the genetic diversity of the 11 different Tibetan sheep populations during the process of phylogenetic evolution, Pi and Tajima's D tests were conducted.

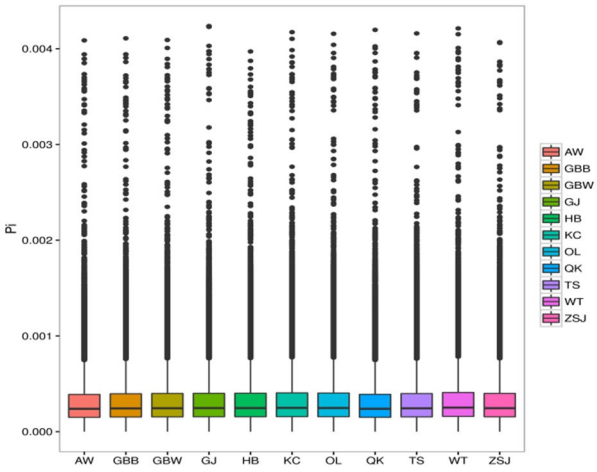

Fig. 7 Nucleic acid diversity results of Tibetan sheep

The Pi values of the 11 Tibetan sheep populations range from 0.0001 to 0.005, indicating that these populations may have experienced certain selection pressures or genetic drift events, leading to genomic loci being relatively conserved within the populations with low levels of variability and reduced genetic diversity (Fig. 7).

For most chromosomal segments of the 11 Tibetan sheep populations, Tajima's D values are greater than 0, indicating that the number of haplotypes exceeds the number of polymorphic sites, suggesting the presence of numerous alleles at intermediate frequencies and rare alleles existing at low frequencies. This implies a scarcity of rare alleles within the populations, with neutral alleles predominating. The populations are in a state of balancing selection, where the population size and density are maintained in dynamic equilibrium over a certain period (Fig. 8).

The F_{ST} values between the various populations range from 0.02 to 0.07, indicating a low to moderate level of genetic differentiation. Specifically, the F_{ST} values between AW and GBB, GBW, QK, and TS; GBB and QK, TS; GBW and QK, TS; QK and TS, ZSJ fall within the range of 0.05 to 0.15, representing a moderate level of genetic differentiation. In contrast, the F_{ST} values between the remaining populations are all less than 0.05, indicating a low level of genetic differentiation (Fig. 9).

Discussion

The natural evolution and frequent human activities directly influence the population distribution and dynamics of species. Intense human intervention disrupts interactions between populations and intraspecific selection. Directional selection for economically favorable traits in individuals can increase selection pressure, leading to frequent instances of inbreeding within populations and

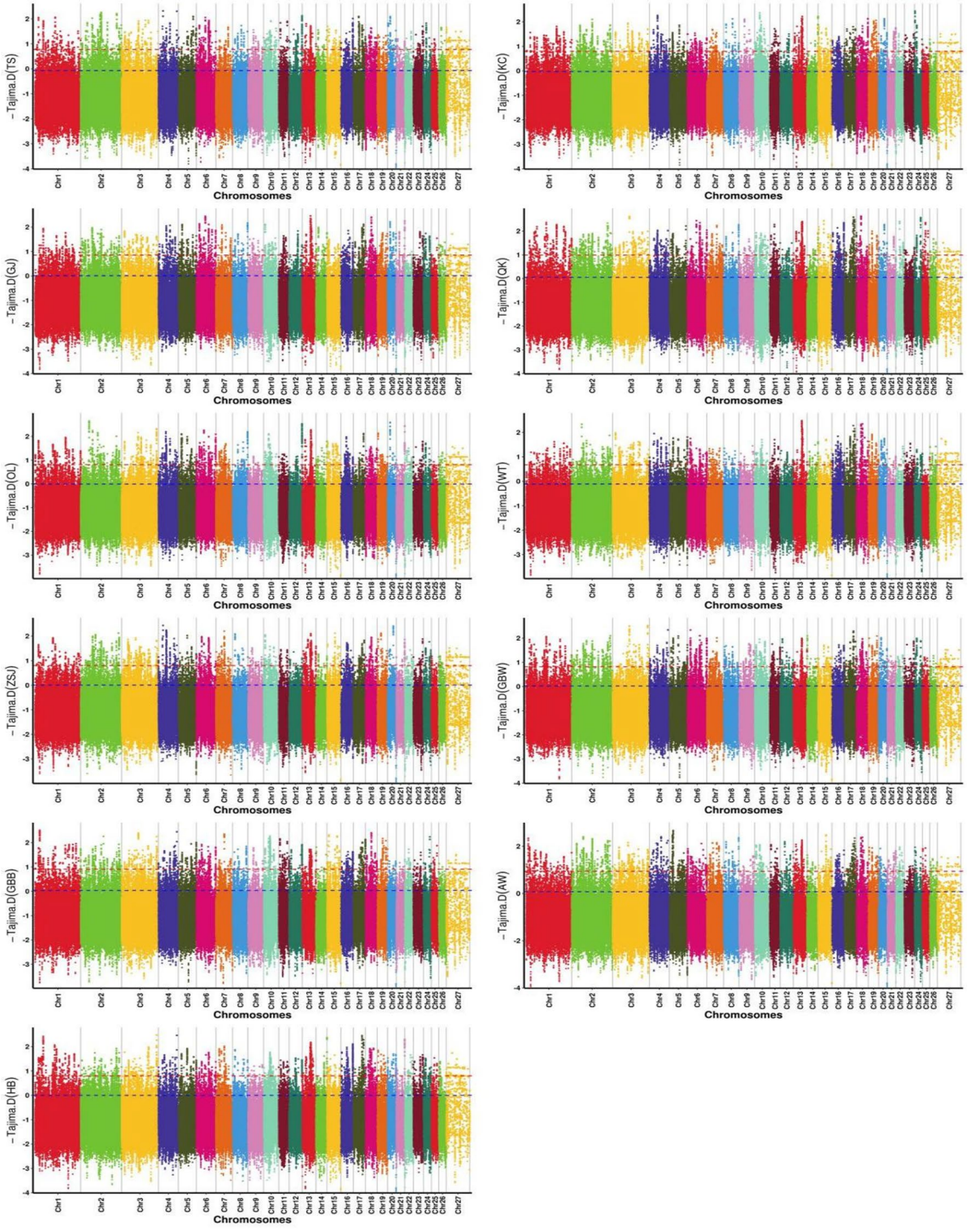

Fig. 8 Tajima's D analysis results of Tibetan sheep population

resulting in reduced genetic heterozygosity in the population's genome [23].

From the PCA and genetic structure analysis of the 11 Tibetan sheep populations, a certain level of shared genetic ancestry was observed between the Tibetan sheep populations ZSJ and WT from Qinghai Province, and OL, GJ, KC, and TS from Gansu Province, providing indirect evidence for a consistent origin of the Tibetan sheep populations. The close geographical proximity, similar genetic background, and common breeding practices among the Tibetan sheep populations have led to the manifestation of similar trait characteristics and

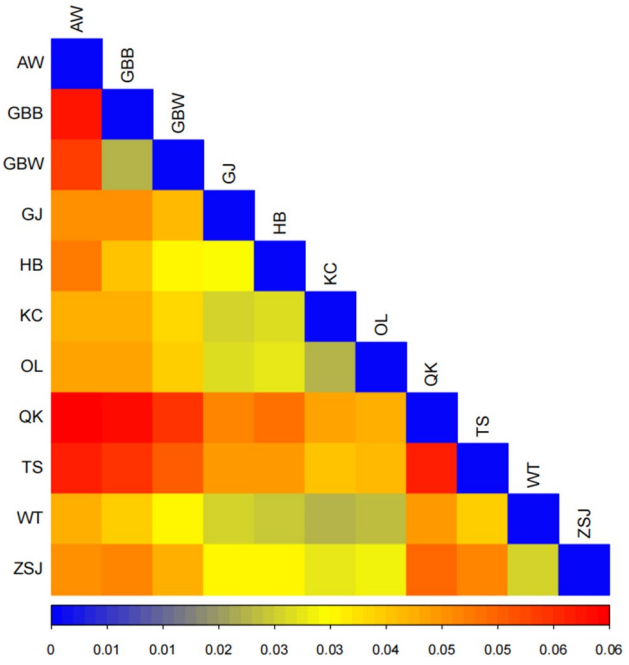

Fig. 9 Genetic differentiation analysis of Tibetan sheep population

subsequently resulted in a low level of genetic differentiation among these populations. This finding is in line with the results of Agraw where low pairwise genetic differentiation indices were observed among the six subpopulations of Sudanese sheep, possibly attributed to the consistent tail morphology types within these Sudanese sheep subpopulations [24]. Furthermore, the current population structure patterns of Tibetan sheep populations may also be related to early human historical activities. Human behavioral activities have to some extent influenced the historical dynamics of Tibetan sheep populations. Migration, admixture, assimilation, and expansion of early human populations can impact the population size and distribution of Tibetan sheep [9]. As early as 40,000 years ago, there were traces of human activities on the Qinghai-Tibet Plateau. In the late Paleolithic period, the Nwya Devu culture emerged in East Asia on the Qinghai-Tibet Plateau. With human settlement and the spread of Paleolithic farming tools and domesticated crops on the Qinghai-Tibet Plateau, there was a short-term population expansion, leading to the expansion of domesticated livestock, a trend supported by the historical dynamics of Tibetan sheep populations. Between 40,000 and 20,000 years ago, there was a population expansion within Tibetan sheep populations [25, 26]. The Tibetan sheep populations migrated along river valleys with human populations, experiencing continuous movements upstream and downstream. During the migration towards the Qinghai-Tibet Plateau, humans migrated to lower altitudes (below 3000 m) between 30,000 and 20,000 years ago, to higher altitudes (3000–4000 m) around 15,000 years ago, and eventually reached extremely high altitudes (above 4000 m) after 6,000 years [27]. Throughout this process, humans continuously migrated, assimilated, expanded, and formed stable tribes, while Tibetan sheep also migrated, dispersed, interbred, and fused into stability [28]. This led to a certain level of genetic admixture between different Tibetan sheep populations, resulting in a low level of genetic differentiation. The phylogenetic tree results also support this hypothesis. Among the 11 Tibetan sheep populations, WT and the outgroup exhibit a close genetic relationship, with WT representing the most ancestral lineage among the Tibetan sheep populations and retaining more ancestral characteristics. The AW and ZSJ populations subsequently evolved. Following this, the populations from Gansu, including GJ, KC, QK, and OL, clustered together and gradually diverged, with the TS population from Gansu evolving and separating afterward. The Tibetan sheep populations from Tibet, namely GBB, GBW, and HB, diverged last. An interesting finding is that although both ZSJ and WT inhabit high-altitude regions in Qinghai Province and migrated to these high-altitude areas approximately 15,000 years ago, ZSJ shows a closer genetic relationship to AW compared to WT, likely due to their geographical proximity [27]. This result indicates that the formation of Tibetan sheep population structure is influenced by multiple factors. Discrepancies with the results of Yang et al. may stem from differences in research methods, population sizes, and total population numbers [29].

Populations with higher genomic genetic diversity often exhibit better environmental adaptability, population resilience to disturbances, and evolutionary potential [30]. The results of this study indicate that the Tajima's D values of all 11 Tibetan sheep populations are greater than 0, suggesting recent population contraction or balancing selection in Tibetan sheep populations. Research has shown that long-term artificial selection in sheep may decrease genomic diversity [31, 32]. This is consistent with the nucleotide diversity results of Tibetan sheep genomes. Another possible reason for this phenomenon may be related to population breeding management practices. Tibetan sheep are raised through grazing, and confinement in closed environments can lead to increased levels of inbreeding and reduced genomic heterozygosity, ultimately resulting in decreased population diversity [24, 33, 34].

Genomic variation is the foundation of biodiversity, with genetic variation representing the common, non-random genetic changes that occur among individuals or populations during the evolutionary process, enriching genomic diversity [35]. Genomic variation drives species evolution, as mutations continually propel species to evolve. Stable inheritance of mutations to offspring

results in genomic diversity among the offspring, enhancing population adaptability and enabling species to thrive in changing environments. The identification of potentially beneficial variant sites is of significant importance for population genetic improvement and genetic diversity. The Tibetan sheep genome exhibits a high level of SNP variation, and numerous studies have demonstrated that genetic variations within sheep genomes can significantly impact environmental adaptability, reproductive traits, growth characteristics, fat deposition, and heat stress responses [36–40]. Based on this, this study investigated the genetic differences among 11 Tibetan sheep populations resulting from genetic genomic variation in the breeding process, which can lead to genetic differentiation among populations. The level of genetic differentiation between populations was assessed, with AW showing moderate genetic differentiation levels with GBB, GBW, QK, TS; GBB showing moderate genetic differentiation levels with QK, TS; GBW showing moderate genetic differentiation levels with QK, TS; QK showing moderate genetic differentiation levels with TS, ZSJ, while the remaining populations exhibited low genetic differentiation levels. Combining the genetic differentiation results with geographical data, the genetic differentiation among sheep populations may be related to environmental adaptation, geographical distribution, and population type [24, 41]. The relatively low genetic differentiation indices among the genomes of different populations may be attributed to similar environmental conditions and altitudes, resulting in minimal differences among populations.

Population genetic diversity can to some extent reflect a species' ability to adapt to changes. Higher levels of population genetic diversity can enhance the survival chances of individuals in changing environments, while lower levels of population genetic diversity may lead to local extinctions of populations [30, 42]. Population genetic diversity is influenced by multiple factors. Inadequately scientific and rational breeding mating practices can lead to increased inbreeding frequency, higher genomic selfing rates, and consequently reduced genetic diversity [43]. Effective population size also impacts the level of genetic diversity, as smaller populations are more prone to inbreeding depression, potentially even facing extinction risks that result in decreased or lost genetic diversity. Additionally, genetic drift, natural selection, and other factors can also affect the level of genetic diversity [44, 45]. The results of this study indicate that the 11 Tibetan sheep populations exhibit varying degrees of low genomic diversity levels, emphasizing the urgent need for the implementation of conservation measures for Tibetan sheep populations. Establishing live gene banks and optimizing breeding programs can help enhance population diversity to some extent. For populations with lower diversity levels such as TS, WT, and GBB, preserving semen or embryos through cryopreservation can help reduce the frequency of inbreeding while setting up conservation facilities. With advancements in biological technologies such as DNA preservation, oocyte cryopreservation, cloning, cryopreservation of embryonic stem cells, and increasing research on genetic mapping and gene isolation and cloning related to desirable traits, future research and practices focused on the conservation of Tibetan sheep diversity will need to be further deepened.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10800-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Acknowledgements

The authors would like to express their gratitude to all reviewers who participated in the review of the paper, as well as to GENE DENOVE, which provided sequencing-related services for the trial.

Author contributions

JBL and ZKL participated in the experiment funding application, experimental design, and article review. LXS participated in the experimental design, operation, data processing, and article writing. CY and TTG participated in experiment guidance, procurement of experimental materials. MSZ and YQB participated in part of the analysis of the trial. All authors reviewed the manuscript.

Funding

This study was supported by the National Key R&D Program of China (2021YFD1600703), the Innovation Project of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2015-LIHPS), the Science and Technology Program of Gansu Province (22CX8NA014), the National Technical System for Wool Sheep Industry (CARS-39-02), Research on the Excavation, Protection and Utilization of New Genetic Resources of Livestock and Poultry in Gansu Province (GNKJ-2021-36).

Data availability

Raw FASTQ files for whole-genome sequencing were deposited in the NCBI Sequence Read Archive (SRA) and have been assigned BioProject accession number: PRJNA1138910.

Declarations

Ethics approval and consent to participate

All procedures in this experiment were conducted in accordance with the regulations of the Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences. Sample collection and experimentation were fully compliant with the requirements of the China Animal Welfare Committee and approved by the Animal Ethics Committee of the Lanzhou Institute of Husbandry and Pharmaceutical Sciences of CAAS (Ethic approval file No. NKMYD201805).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 9 July 2024 / Accepted: 13 September 2024
Published online: 30 September 2024

References

1. A ZM, B H. The Initial Domestication of Goats (Capra hircus) in the Zagros Mountains 10,000 Years Ago. Science (New York, NY). 2000;287(5461):2254-7.
2. Wang YF. Preliminary Study of Molecular Archaeology of Domestic Sheep Duringthe Longshan Period in Northwest China. In.; 2023.
3. Zuo HR. Review and Prospect of the animal archaeology of the domestic sheep/goat in China. Cult Relics South China. 2017;(01):155–63.
4. Jing XP, Wang WJ, Degen AA, Guo YM, Kang JP, Liu PP, et al. Small intestinal morphology and sugar transporters expression when consuming diets of different energy levels: comparison between tibetan and small-tailed Han sheep. Animal. 2022;16(3):100463. https://doi.org/10.1016/j.animal.2022.100463.
5. Yao WL, Sha YZ, Guo XY, Pu XN, Xu Y, Wang JQ, et al. Analysis of Meat Quality, Nutritional Components and expression characteristics of Meat Quality-related genes in Tibetan Sheep at different altitudes. J Agricultural Sci Technol. 2024;26(03):66–75. https://doi.org/10.13304/j.nykjdb.2022.0385.
6. Zheng H, Zhao T, Xu YC, Zhang DG, Song YF, Tan XY. Dietary choline prevents high fat-induced disorder of hepatic cholesterol metabolism through SREBP-2/HNF-4α/CYP7A1 pathway in a freshwater teleost yellow catfish Pelteobagrus fulvidraco. Biochim Biophys Acta Gene Regul Mech. 2022;1865(7):194874. https://doi.org/10.1016/j.bbagrm.2022.194874.
7. Jiao J, Wang T, Zhou J, Degen AA, Gou N, Li S, et al. Carcass parameters and meat quality of tibetan sheep and small-tailed Han sheep consuming diets of low-protein content and different energy yields. J Anim Physiol Anim Nutr (Berl). 2020;104(4):1010–23. https://doi.org/10.1111/jpn.13298.
8. Han B, Tian D, Li X, Liu S, Tian F, Liu D, et al. Multiomics analyses provide New Insight into Genetic Variation of Reproductive adaptability in Tibetan Sheep. Mol Biol Evol. 2024;41(3). https://doi.org/10.1093/molbev/msae058.
9. Hu XJ, Yang J, Xie XL, Lv FH, Cao YH, Li WR, et al. The Genome Landscape of Tibetan Sheep reveals adaptive introgression from Argali and the history of early human settlements on the Qinghai-Tibetan Plateau. Mol Biol Evol. 2019;36(2):283–303. https://doi.org/10.1093/molbev/msy208.
10. Hong C, Zhuangbiao Z, Jiayue W, Rasmus ALJ, Yudong H. Long divergent haplotypes introgressed from wild sheep are associated with distinct morphological and adaptive characteristics in domestic sheep. PLoS Genet. 2023;19(2):e1010615. https://doi.org/10.1371/journal.pgen.1010615.
11. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.
12. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8. https://doi.org/10.1038/ng.806.
13. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033.
14. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. https://doi.org/10.1093/nar/gkq603.
15. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585–95. https://doi.org/10.1093/genetics/123.3.585.
16. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8. https://doi.org/10.1093/bioinformatics/btr330.
17. Nei M. New York Chichester. West Sussex: Columbia University; 1987.
18. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 2017;49(2):303–9. https://doi.org/10.1038/ng.3748.
19. Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P, et al. Convergent genomic signatures of domestication in sheep and goats. Nat Commun. 2018;9(1):813. https://doi.org/10.1038/s41467-018-03206-y.
20. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science. 2019;364(6446). https://doi.org/10.1126/science.aav6202.
21. Lv FH, Cao YH, Liu GJ, Luo LY, Lu R, Liu MJ, et al. Whole-genome resequencing of Worldwide Wild and Domestic Sheep Elucidates Genetic Diversity, Introgression, and agronomically important loci. Mol Biol Evol. 2022;39(2). https://doi.org/10.1093/molbev/msab353.
22. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.
23. Kinghorn B, Kinghorn A. Management of diversity and inbreeding when importing new stock into an inbred population. J Hered. 2023;114(5):492–503. https://doi.org/10.1093/jhered/esad027.
24. Agraw A, Gurja B, Abdulfatai T, Tadelle D, Olivier HMH. Genome-wide genetic diversity and Population structure of local Sudanese Sheep populations revealed by whole-genome sequencing. Diversity. 2022;14(11):895.
25. Xiao G. The earliest evidence of human adaptation to high-altitude extreme environments has been found in the hinterland of the tibetan plateau. Science. 2019;71(01):56.
26. Zhang XL, Ha BB, Wang SJ, Chen ZJ, Ge JY, Long H, et al. The earliest human occupation of the high-altitude Tibetan Plateau 40 thousand to 30 thousand years ago. Science. 2018;362(6418):1049–51. https://doi.org/10.1126/science.aat8824.
27. Zhang DJ, Dong GH, Wang H, Ren XY, Ha BB, Qiang MR, et al. The historical process and possible driving mechanism of prehistoric humans spreading to the Qinghai Tibet Plateau. Scientia Sinica(Terrae). 2016;46(08):1007–23.
28. Sirous E, Amiri RM, Osman KM, Goutam S. Genomic scans for selective sweeps through haplotype homozygosity and allelic fixation in 14 indigenous sheep breeds from Middle East and South Asia. Sci Rep. 2021;11(1):2834.
29. Yang XJ. Stady on genetic diversity and breeds differentiation of tibet sheep. In.; 2001.
30. F FW EHE. Quantifying the extinction vortex. Ecol Lett. 2006;9(1):51–60.
31. Corbett-Detig RB, Hartl DL, Sackton TB. Natural selection constrains neutral diversity across a wide range of species. PLoS Biol. 2015;13(4):e1002112. https://doi.org/10.1371/journal.pbio.1002112.
32. Whitlock MC, Bürger R. Fixation of new mutations in small populations. In: Ferrière R, Dieckmann U, Couvet D, editors. Evolutionary Conservation Biology. Cambridge: Cambridge University Press; 2004. pp. 155–70.
33. Spielman D, Brook BW, Briscoe DA, Frankham R. Does inbreeding and loss of genetic diversity decrease disease resistance? Conserv Genet. 2004;5(4):439–48.
34. Gornas N, Weimann C, Hussien AE, Erhardt G. Genetic characterization of local Sudanese sheep breeds using DNA markers. Small Ruminant Res. 2010;95(1):27–33. https://doi.org/10.1016/j.smallrumres.2010.08.009.
35. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–59. https://doi.org/10.1093/GENETICS/155.2.945.
36. Shadan T, Mohsen G, Ghodrat rM, Hossein MM, Roy C, Stephen M et al. Genome-wide identification of copy number variation and association with fat deposition in thin and fat-tailed sheep breeds. Sci Rep. 2022;12(1).
37. Luo Y, Zhang M, Guo Z, Dwi W, Xu H, Jiang F, et al. Insertion/Deletion (InDel) variants within the Sheep Fat-Deposition-related PDGFD gene strongly affect morphological traits. Animals. 2023;13(9):1485. https://doi.org/10.3390/ani13091485.
38. Naga AMA, Alsamman AM, Allali AE, Elshafie MH, Abdelal ES, Abdelkhalek TM, et al. Genome-wide analysis identified candidate variants and genes associated with heat stress adaptation in Egyptian sheep breeds. Front Genet. 2022;13:898522. https://doi.org/10.3389/fgene.2022.898522.
39. Wen Y, Wang E, Wang X, Qing S, Chaogetu B, Wang C, et al. Copy number variations of LRRFIP1 gene and the relationship with growth traits in four Chinese sheep. Animal Biotechnol. 2022;34(7):3008–15. https://doi.org/10.1080/10495398.2022.2126981.
40. Morteza M, Arsalan B, Zahra R, Farhad G, Amiri RM. Quantifying parent-of-origin variation in growth and reproductive traits of Kermani sheep. J Agricultural Sci. 2022;160(5):391–6. https://doi.org/10.1017/S0021859622000405.
41. Madrigal L, Ware B, Miller R, Saenz G, Chavez M, Dykes D. Ethnicity, gene flow, and population subdivision in Limón, Costa Rica. Am J Phys Anthropol. 2001;114(2):99–108. https://doi.org/10.1002/1096-8644(200102)114:2%3C99::AID-AJPA1010%3E3.0.CO;2-V.

42. Simone C, Vincenzo L, Gabriele S, Salvatore M, Teresa SM, Slim BJ, et al. A comprehensive analysis of the genetic diversity and environmental adaptability in worldwide Merino and Merino-derived sheep breeds. Genet Selection Evol. 2023;55(1):24.
43. Eydivandi S, Sahana G, Momen M, Moradi MH, Schönherz AA. Genetic diversity in Iranian indigenous sheep vis-à-vis selected exogenous sheep breeds and wild mouflon. Anim Genet. 2020;51(5):772–87. https://doi.org/10.1111/age.12985.
44. Lynch M, Conery J, Burger R. Mutation Accumulation and the extinction of small populations. The American naturalist. The American Society of Naturalists; 1995;146(4):489–518.
45. Charlesworth B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009;10(3):195–205. https://doi.org/10.1038/nrg2526.

Publisher's note