Genetic Diversity and Runs of homozygosity analysis of Hetian Sheep Populations Revealed by Illumina OvineSNP50 BeadChip

Zhipeng Han
Tarim University

Wen Zhou
Tarim University

Lulu Zhang
Tarim University

Ruotong Wang
Tarim University

Chunjie Liu
Tarim University

Xinyu Bai
Tarim University

Shudong Liu (✉ liushudong63@126.com)
Tarim University

Research Article

Keywords: Illumina OvineSNP50 BeadChip, Hetian sheep, Population genetic structure, ROH

Posted Date: September 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2022113/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Hetian sheep have a long history, a wide production area and rich genetic resources. Due to the different growing environment and feeding methods, wool, body shape and other traits also have some differences. Hetian sheep can be roughly divided into three groups, but the classification at the genome level is not clear.

Results

We randomly selected 84 healthy, adult ewes were randomly selected from three different ecological regions in Hetian area of Xinjiang, China. DNA was extracted from Ear tissues using phenol chloroform method to produce Illumina OvineSNP50 BeadChip. The SNP chip data were analyzed by Principal component analysis (PCA), neighbor-joining (NJ) phylogenetic and admixture for the population genetic structure of Hetian sheep, and the degree of population linkage was analyzed based on linkage disequilibrium (LD). Finally, the genetic region selection of Hetian sheep genetic germplasm resource populations in three different ecological environments was obtained by runs of homozygosity (ROH) analysis. The obtained candidate genes were annotated with Oar_v4.0 and enriched by GO and KEGG for analysis. Main screened 31 candidate genes adapting to high altitude environment were obtained in the Mountain type. Among them, Genes related to bone cell generation, differentiation and maintenance of bone homeostasis WNT6, WNT10A, CHSY1, etc. Genes related to tooth and tongue development LEF1, TP63, PRDM16, etc. Hearing and vision-related genes RBP4, ATF6, JAG1, etc. Main screened 22 candidate genes related to economic traits were obtained in the Grass type. Among them, Reproduction-related genes PLA2G4F, ACVR1, ADCY2, etc. Growth-related genes CAPN3, YAP1, FGF9, etc.

Conclusions

Hetian sheep can be classified into three subtypes at the genomic level: Mountain, Mountain-Grass and Grass types. The Grass type evolved from the Mountain type, and some of the Grass type further evolved into the Mountain-Grass type, and the genetic relationship between the Mountain-Grass type and Grass type was closest. Genetic structure and ROH analysis of Hetian sheep based on genomic microarray technology revealed the Mountain-Grass type strain. We enriched the genetic diversity of Hetian sheep germplasm resources, provided directions for Hetian sheep breed conservation, and found that genes related to multiparous trait exist in Mountain-Grass and Grass types Hetian sheep, which provides ideas for the selection and breeding of multiparous Hetian sheep.

Background

The main production area of Hetian sheep is in the Hetian region of southern Xinjiang, bordered by the Taklimakan Desert in the north and the northern foothills of the Kunlun Mountains in the south, with little rain and drought throughout the year. Hetian sheep have high adaptability to harsh ecological conditions
and are known for their carpet wool[1]. During the survey, it was found that the overall geographical trend of the Hetian sheep production area shows a special topography of high south and low north. The size and genetic characteristics of their body shape also change due to different growth environments and feeding methods[2]. According to these characteristics, it is of great importance to correctly understand and protect the genetic resources of Hetian sheep.

With the advent of high-density SNP Bead Chip, population genetic structure and analysis of population variability can be more accurately determined[3]. Adam Abied et al. Prepared Ovine Infinium 600K SNP BeadChip by extracting DNA from the blood of 121 animals representing five archetypes of thin-tailed Desert Sheep in Sudan and 65 individuals of four breeds of Chinese sheep. Genomic variation in the African thin-tailed Desert Sheep sampled from Sudan. Was analyzed at the genomic level and found that it had high levels of genetic diversity, but low levels of genetic differentiation. Genes such as PDGFRA and CAMK4 associated with parasite resistance, coat color, differential fat deposition and tail morphology, and genes such as SOCS2 and AP1G1 associated with production, reproduction and multiple birth traits were also identified. Providing a basis for understanding the genome structure of African native sheep[4]; Zewdu Edea et al. Explored the genetic diversity and population structure of Ethiopian sheep populations using a high-density 600K SNP BeadChip of 72 Ethiopian sheep from 5 species. It was determined that Ethiopian sheep populations are roughly clustered according to their geographic distribution and tail phenotype, while also elucidating the evolutionary history of sheep on the African continent[5]; Tatiana E. Deniskova et al. Used genome-wide Ovine SNP50 BeadChip data to explore the population structure of 396 local Russian sheep of 25 species and determined that Russian coarse wool is primarily identified by tail type and that local fat-tailed coarse wool breeds and Southwest Asian (Iranian) sheep share a common ancestor[6]; Jie Cheng et al. Analyzed 10 Chaka sheep and 30 Bayinbuluke, Tan and Oula sheep based on sheep genome resequencing, and found that Chaka sheep were significantly differentiated from Bayinbuluke, Tan and Oula sheep, and Chaka sheep had an independent ancestor. In addition, neurofibromin 1 (NF1) and myomesin 1 (MYOM1) genes related to muscle development were identified in Chaka sheep by Fst and XP-CLR genome-wide scanning[7]; Adam Abied et al. Analyzed the genomic diversity, population structure and genomic selection of 5 Chinese native sheep at the genomic level using a 600K SNP chip with five 96 Chinese native sheep markers. The results showed that the genetic variability was between low and medium[8]; M. A. Stoffel et al. Explored the genetic structure of inbreeding depression in 5952 wild Soay sheep using the ROH length classes and found that a 10 % increase in the inbreeding rate resulted in a 60 % decrease in the survival rate of the offspring, that the degree of inbreeding depression, decreased with age, and that many alleles with weakly deleterious effects or at low frequencies probably contribute to inbreeding depression in survival[9]; E. F. Dzomba et al. Explored the number and distribution of ROH from Illumina OvineSNP50 BeadChip data of 400 native South African sheep breeds from 14 species and 623 African, Asian and European sheep breeds markers from 17 species. They found that the length of ROH detected in all breeds was 1-6Mb, accounting for 88.2 %, and the ROH was almost no more than 48Mb. At the same time, metabolism and immune response trait genes ST3GAL3, ADRA2C, etc. and angular growth gene RXFP2 were also found in the ROH fragment[10].
In this study, Illumina Ovine SNP50 BeadChip was used to analyze the genetic structure of germplasm resources in 3 groups of Hetian sheep, and the Mountain-Grass type of Hetian sheep was found. Based on ROH analysis, it was found that there were genes related to prolificacy traits in Grass type Hetian sheep, which provided theoretical support for Hetian sheep protection and germplasm resource utilization.

**Methods**

**Hetian sheep samples**

In this experiment, 84 healthy adult female Hetian sheep with the same age and in good condition were randomly selected from three different ecological environment areas in the Hetian area of Xinjiang in 2018, and ear tissues were collected (Fig.1). There were 32 Hetian sheep of Mountain type in Minfeng County (hereafter referred to as MTS), 23 Hetian sheep of Mountain-Grass type in Yutian County (hereafter referred to as MGTS) and 29 Hetian sheep of Grass type in Moyu County (hereafter referred to as GTS).

**DNA extraction & SNP microarray preparation**

After extracting DNA from the collected Hetian sheep ear tissue samples using the phenol-chloroform method\[11\], Illumina OvineSNP50 BeadChip was prepared and the formed data signal was processed and typed for data visualization using Genome Studio software.

**SNP quality control**

Genotype quality control using Plink1.07\[12\], QC criteria: (1) Rejection detection rate < 0.90 (2) Minor allele frequency (MAF) < 0.05 within the breeds;(3) p-value for Hardy-Weinberg equilibrium < 0.00001.

**Principal coordinates analysis**

To study the genetic relationship between individuals and populations, PCA was performed on 46329 eligible SNP locis in 84 samples using the make-grm and pca commands in PLINK1.07\[13,14\]. Then use R software to draw a two-dimensional plan between PC1, PC2 and PC3, and analyze the clustering between different individuals in the population.

**Phylogenetic analysis**

Identity By State (IBS) matrix was constructed using PLINK 1.07. The matrix quantified the genetic similarity between individuals and was used to measure the pairwise genetic distance between varieties\[15\]. Then, based on the SNP, the NJ phylogenetic tree was established by the neighbor-joining method of the Mega 6.0 software.

**Admixture analysis**
Under the assumptions of Hardy-Weinberg equilibrium, complete linkage equilibrium, and no prior information about the population of individual origin, K values (assuming the number of ancestral populations) of 2-4 were analyzed using ADMIXTURE software to determine the ancestors to infer the origin of the species and quantify the admixture value of the population\[16\]. Apply their respective Cross Validation (CV) errors.

**LD analysis**

This method uses $r^2$ as a measure of LD. $r^2$ represents the degree of statistical and genetic correlation between the two loci ($0 < r^2 < 1$), which is stable and insensitive to changes in gene frequency\[17-19\]. PopLDdecay software was used to detect the level of LD and LD attenuation of SNP loci in 3 populations in the range of 100Kb, and then perlPlot _ OnePop.pl command was used to draw LD attenuation images according to the statistical results.

**ROH analysis**

ROH analysis was performed on the three populations using the homozyg command in PLINK1.07 software. The parameters are as follows: 1) a maximum of one missing and one heterozygous SNP were allowed within an ROH and SNP window; 2) The number of SNPs constituting a single ROH cannot be less than 30; 3) The minimum length of an ROH was 1000 Kb; 4) the maximum gap within an ROH was 250 Kb; 5) the minimum marker density was set to one SNP every 100 kb; other parameters are default.

By classifying the length of ROH fragments to compare the distribution of ROH in each population, they can be divided into 3 groups: 0-5 Mb, 5-10 Mb and $\geq$ 10 Mb. (1) ROH coverage (ROH%) refers to the proportion of ROH on CHR; (2) The percentage of the longest ROH on CHR (MROH%).

**Genome region selection& Candidate gene identification**

The distribution of ROH in the CHR of Hetian sheep in the 3 regions was counted, and only the overlapping ROH fragments of MTS and MGTS, GTS and MGTS were used as the genomic segments of MTS and GTS that can be inherited to MGTS\[9\]. Annotated in the sheep genome Oar _ v4.0 based on the physical location of the heritable segments. DAVID (https://david.ncifcrf.gov/) was used for cluster analysis of genes, and Quick GO (https://www.ebi.ac.uk/QuickGO/) and KEGG (https://www.genome.jp/kegg/kegg2.html) were used to query the relevant information and functions of annotated genes\[20\]. Finally, candidate genes related to environmental adaptation were screened according to the NCBI (https://www.ncbi.nlm.nih.gov/pmc/) literature search.

**Results**

Population genetic structure
The first three main components explained 56.24%, 24.79% and 18.97% of the total variation. PC 1 and PC 2 were plotted for the X and Y axes, and the three populations were clustered into three groups (Fig.2a: Additional file 1); PC1 and PC3 were drawn scatter plots for the X and Y axes, GTS and MGTS were clustered on the upper right side of the graph and the two populations were very close (Fig.2b: Additional file 1); PC2 and PC3 draw scatter plots for the X and Y axis. MGTS, GTS and MTS populations are close to each other and inseparable (Fig.2c: Additional file 1).

Hetian sheep populations in three different growing environments can be divided into three branches. MTS is clustered into one branch, MGTS and GTS are clustered into another branch, and MTS and GTS branches are far from each other (Fig.3).

Each column in the figure represents a sample. The 0.0~ 1.0 scales on the left represents the proportion of specific genes. Righter K assumes 2~ 4 ancestral populations, of which the best K value is 3. At the level of K = 2 ~ 3, MGTS and GTS showed a common ancestor and MGTS was bred from GTS. When K = 3 ~ 4, the K values of MGTS and GTS groups did not change, and MTS had two ancestors (Fig.4: Additional file 2).

Linkage disequilibrium

The LD of the 3 populations showed a gradual decline. And \( r^2 \) decreased rapidly when the physical distance of the SNP locus increased. Among them, MTS decays fastest, followed by GTS, and finally MGTS; among the attenuation degrees of the 3 populations, the LD level of MTS was the lowest after 80kb, that of GTS was in the middle and that of MGTS was the highest (Fig.5).

Runs of Homozygosity

A total of 700 ROH were detected in 3 populations. 312 ROH were detected in MTS, 207 in 0-5Mb, 79 in 5-10Mb and 26 in \( \geq 10Mb \). 303 ROH were detected in GTS, 213 in 0-5Mb, 67 in 5-10Mb and 23 in \( \geq 10Mb \). 85 ROH were detected in MGTS, 68 in 0-5Mb, 14 in 5-10Mb and 3 in \( \geq 10Mb \). The length of ROH is mainly concentrated within 10 Mb (Fig.6: Additional file 3, Additional file 4, Additional file 5).

The ROH of the 3 groups was grouped according to 0 ~ 5Mb, 5 ~ 10Mb, \( \geq 10Mb \), and the number of ROH on each CHR and the proportion of the longest ROH on the CHR of the 3 groups were studied. The ROH % and MROH% of MTS and GTS were higher than those of MGTS and MGTS. MGTS had the least ROH number and the shorter ROH length. No ROH was detected on CHR 11, 20, 21, 23 and 24 (Fig.7: Additional file 3, Additional file 4, Additional file 5).

Identification of ROH Islands and Gene Annotation

Annotation of the complete genome heritable region Oar_v4.0 in three types of Hetian sheep and 8660 specific genes were detected in MTS, 1937 specific genes were detected in MGTS, and 6427 specific genes were detected in GTS. After the intersection of MTS and MGTS, the specific genes shared by GTS were removed, and 498 candidate genes specific to MTS and MGTS were remaining. After the
intersection of GTS and MGTS, the specific genes shared by MTS were removed, and 191 candidate genes specific to the GTS and MGTS were remaining (Fig.8: Additional file 6).

GO and KEGG enrichment were performed on 498 candidate genes unique to the MTS and MGTS. Through GO and KEGG enrichment, a total of 407 MGTS and MTS-specific genes were screened and enriched to 77 pathways. The selected gene CC was mainly enriched into three pathways: GO: 0005634 ~ nucleus, GO: 0005829 ~ cytosol and GO: 0005737 ~ cytoplasm, with a total of 125 specific genes. BP was mainly enriched in 10 pathways: GO: 0034765 ~ regulation of ion transmembrane transport, GO: 0016055 ~ Wnt signaling pathway, GO: 0007605 ~ sensory perception of sound, GO: 0042475 ~ odontogenesis of dentin-containing tooth, GO: 0043586 ~ tongue development, GO: 0007162 ~ negative regulation of cell adhesion, GO: 0007162 ~ negative regulation of cell adhesion, GO: 0001654 ~ eye development, GO: 0042491 ~ auditory receptor cell differentiation. There are mainly 31 significant genes; Four significant genes WNT6, WNT10A, LEF1, TCF7L1 in the KEGG _ PATHWAY were simultaneously enriched in the pathways of oas04310: Wnt signaling pathway, oas04916: Melanogenesis, oas05217: Basal cell carcinoma. These genes are mainly significantly associated with the phenotypes of neurodevelopment, bone morphogenesis, tooth production, tongue development, vision, and hearing (Fig.9: Additional file 7; Fig.10: Additional file 8; Tab.1: Additional file 7, Additional file 8).

Determining the genetic changes behind these phenotypic expressions helps to elucidate the genetic mechanism of Hetian sheep under natural conditions and provide corresponding breeding strategies for resisting harsh environments.

GO and KEGG enrichment were performed on 191 candidate genes specific to the GTS and MGTS. Through GO and KEGG enrichment, a total of 160 MGTS and GTS-specific genes were screened and enriched to 52 pathways. The selected gene CC was mainly enriched in GO: 0005829 ~ cytosol pathway, and there were 27 specific genes. BP was mainly enriched in 10 pathways: GO: 0003143 ~ embryonic heart tube morphogenesis, GO: 0045669 ~ positive regulation of osteoblast differentiation, GO: 0000082 ~ G1 / S transition of mitotic cell cycle, GO: 000902 ~ cell morphogenesis, GO: 0007179 ~ transforming growth factor β receptor signaling pathway, GO: 0048333 ~ mesodermal cell differentiation, GO: 0014850 ~ response to muscle activity, GO: 0043149 ~ stress fiber assembly, GO : 0006957 ~ complement activation. There are mainly 13 significant genes; The 10 pathways in the KEGG _ PATHWAY were oas04921 ~ Oxytocin signaling pathway, oas04935 ~ Growth hormone synthesis, oas04810 ~ Regulation of actin cytoskeleton, oas04151 ~ PI3K-Akt signaling pathway, oas04611 ~ Platelet activation, oas05410 ~ Hypertrophic cardiomyopathy, oas05414 ~ Dilated cardiomyopathy, oas04927 ~ Cortisol synthesis and secretion, oas04913 ~ Ovarian steroidogenesis, oas04550 ~ Signaling pathways regulating pluripotency of stem cells. There are mainly 16 significant genes (Fig.11: Additional file 9; Fig.12: Additional file 10; Tab.2: Additional file 9, Additional file 10).

These genes are enriched in the pathways of reproduction, growth and immunity. Identifying the genetic changes behind these phenotypic changes will help to elucidate the genetic mechanism associated with
multiple birth traits after breeding, and provide certain breeding strategies for improving their productivity.

Discussion

Correlation of population genetic structure

Based on SNP50 BeadChip, this study analyzed the SNP loci in the complete genome of Hetian sheep. PC2 V PC3 brings 3 populations aggregation at the same level and indivisible, which means that the 3 populations have the same pedigree, that is, have the same ancestor. The phylogenetic tree shows that the 3 populations have the same 'root node', indicating that the 3 populations have a close genetic relationship. The results were consistent with the results of principal component analysis, indicating that the 3 populations had the small genetic distance and greater genetic consistency. These experimental results are consistent with the actual genetic background of Hetian sheep, which proves that the experimental strategy is reasonable and feasible, and also has a certain control effect in the experiment.

Diversity of population genetic structure

Based on PCA, in the PC1 and PC3 clustering maps, GTS and MGTS were clustered on the right side of the Y-axis, and the 2 groups were closely related. Under normal circumstances, the distribution of clusters is closely related to the geographical distribution of varieties. This result is consistent with the results of the phylogenetic tree, that is, MGTS and GTS are clustered together and have the same 'root node', indicating that GTS and MGTS have the same ancestor.

In addition, Admixture analysis of 84 samples showed that MGTS was cultivated from GTS when K=2-3, which was consistent with the above results. At the same time, it can be speculated that MGTS type is a new strain evolved based on the GTS combined with artificial selection or natural selection. On the other hand, in the PC1 and PC2 clustering maps, the 3 sheep strains are obviously separated. Three Strains of Sheep in the Phylogenetic Tree Divided into 3 Branches. When Admixture analysis K=4, there is no gene doping in MGTS and GTS. In summary\textsuperscript{[21]}, Hetian sheep can be divided into 3 types: MTS, GTS and MGTS at the genomic level.

Population linkage genetic degree

Overall, the 3 types of Hetian sheep can be clearly distinguished and clustered according to geographical distribution. In LD, when the genetic diversity of the population decreases and the linkage between loci increases, the decay rate of LD is the smallest\textsuperscript{[22]}. From the perspective of the attenuation degree of LD, the minimum attenuation distance of MTS is the maximum attenuation rate, which may be due to the high genetic diversity within the population, representing the highest autogenous variation within the population, that is, the low degree of domestication selection. Secondly, compared with GTS and MGTS, MGTS has a further degree of evolution. The maximum MGTS attenuation distance may be due to the high selection caused by
artificial domestication. The results showed that MTS had a closer relationship with the ancestors of Hetian sheep. MGTS had the farthest genetic relationship with the ancestors of Hetian sheep. The genetic relationship between GTS and Hetian sheep ancestors was between MTS and MGTS. Based on the results, it can be inferred that GTS evolved from MTS. At the same time, with the development of agriculture in Grass District, the amount of fertilizer used increased, and the method of shifting sheep downhill by stages to produce manure appeared, which provided a basis for MTS to evolve into the GTS.

Gene region selection situation

ROH are contiguous homozygous segments of the genome and is widely present in the population. Because the gene segment is transmitted directly from the parent to the offspring, the length and number of ROH segments can directly reflect the population genetic history of the population\cite{23,24}. In addition, the long ROH fragment can directly indicate the high degree of inbreeding in the population, that is, the population is close to the ancestor, because the long ROH fragment indicates that the CHR in the population have not undergone multi-generation CHR re-segmentation. In contrast, if the genetic relationship is relatively far, the longest ROH fragments within the population will be very few\cite{25}. In this study, the number of ROH fragments with a length of 0~5Mb in MTS was lower than that in GTS, but the number of ROH fragments greater than 5Mb was higher than that in GTS, indicating that MTS had a closer period of intra-group inbreeding with MTS, that is, MTS had a closer relationship with the ancestors of Hetian sheep; MGTS was much lower than MTS and GTS in the number of ROH fragments and the length of ROH fragments, indicating that MGTS had a longer period of intra-group inbreeding, that is, MGTS has a distant relationship with the ancestors of Hetian sheep, which was consistent with the results of LD. MTS has evolved from mountain to grassland to form GTS, and MGTS has a farther biological evolutionary process, which is a new strain formed by GTS artificial breeding or natural selection.

Exploring the regulatory mechanisms of high-altitude adaptive pathways

For further study the adaptation mechanism of MGTS and MTS performance characteristics significantly expressed by natural selection to the high-altitude environment. In this study, a total of 408 MGTS and MTS specific genes were screened out by GO and KEGG enrichment, which were enriched to 75 pathways. These genes were not expressed in GTS. The results of bioaccumulation clarified the causal relationship between the changes in gene expression and the changes of physiological traits related to performance, which was the basis of adaptive changes of physiological traits in the high-elevation environment.

GO:0034765 (regulation of ion transmembrane transport) mainly include 10 genes, Including the Inwardly rectifying potassium channel related genes are \textit{KCNJ9, KCNJ10, KCNJ11}. Voltage-gated potassium channel related genes are \textit{KCNA2, KCNA10, KCNC1, KCNC2, KCNC4, KCNV1, KCNV2}. GO:0034765 is the basis of neural-electrophysiology. Notably, potassium channels play an important role in neuronal excitability and are associated with inward negative resting membrane potential\cite{26}. It also plays a crucial role in the expression of cells in the central nervous system, spinal cord, kidney, heart, skeletal muscle,
etc\textsuperscript{[27]}. These genes provide physiological guarantee for Hetian sheep to receive various information of internal and external environment in high altitude environment, and then control and regulate the activities of various systems and organs of the body, coordination of movement, and transmission of information through peripheral nerves to maintain the relative balance between the body and the internal and external environment. The genes \textit{WNT6, WNT10A, FRAT2, SFRP5, LEF1, RSPO2, LRP4} enriched in GO:0016055 (Wnt signaling pathway), and are closely related to cells proliferation and differentiation and are involved in the bone cell generation, proliferation and differentiation, embryonic development, organ formation, hair and tooth formation and regeneration, and also play a vital role in wound healing and cardiovascular system. \textit{WNT6} plays an essential role in the normal development process, not only in the embryonic morphogenesis, decidualization, tooth formation and development, maintenance of the normal functioning of Mesenchymal stem cells, but also plays an important role in post-natal homeostasis\textsuperscript{[28]}. \textit{WNT10A} not only signaling can play a pivotal in vivo role in wound healing by regulating the expression and synthesis of collagen\textsuperscript{[29]}, but also plays a vital role in the development and regulation of ectodermal derivatives, which in turn affect the normal formation and regeneration of hair, teeth, skin and nails\textsuperscript{[30]}, while the synergistic effect of \textit{WNT6} and \textit{WNT10A} also promotes osteoblast formation and mutual indemnification\textsuperscript{[31]}. \textit{FRAT2} promotes skeletogeny by affecting Wnt signaling to regulate osteoblast proliferation and differentiation\textsuperscript{[32]}. \textit{SFRP5} is not only highly expressed in retinal pigment epithelial cells and white adipose tissue, but also biologically regulates cell polarity and organ formation during embryonic development\textsuperscript{[33]}. \textit{LEF1} is an important member of the effector protein family LEF and plays a crucial regulatory role in skeletal development, and the \textit{LEF1} transcription factor plays a crucial role in normal osteogenesis and bone loss of \textit{LEF1}\textsuperscript{[34]}. \textit{RSPO2} not only regulates osteoblast activity and facial morphogenesis, but also controls limb development\textsuperscript{[35]}. \textit{LRP4} not only controls the number, morphology and growth of teeth, but also plays a crucial role in maintaining bone homeostasis\textsuperscript{[36,37]}. GO:0030279 (negative regulation of ossification) was mainly included with \textit{CHSY1, MDK, LRP4}. \textit{CHSY1} is essential for skeletal development and increased bone length\textsuperscript{[38]}. \textit{MDK} and \textit{LEF1} play an important role in cartilage development, skeletal development, osteoblast differentiation and regeneration\textsuperscript{[39,40]}. The major genes \textit{WNT6, LEF1, LRP4} and \textit{TP63} enriched in GO:0042475 (odontogenesis of dentin-containing tooth) play an important role in the differentiation of dental cells, skeletal development and osteoblast differentiation\textsuperscript{[28,34,36,41]}. GO:0043586 (tongue development) was mainly enriched with \textit{WNT10A, LEF1, PRDM16}. \textit{PRDM16} is required for normal palatogenesis and play critical and widespread roles during embryonic development\textsuperscript{[42]}. The expression of these genes can be explained by the fact that in the high-altitude environment, where vegetation is sparse, food shortage, and water source is less. Compared to sheep raised on the plain, the Hetian sheep in the high-altitude environment must find food and water to meet their physiological needs if they want to survive. Therefore, they must possess a good exercise capacity. The expression of genes such as \textit{WNT6} and \textit{WNT10A} provides a basis for the formation, proliferation and differentiation of osteocytes and the formation of hair. \textit{TCFL1} and other genes promote the diverse expression of cardiovascular system and play an important role in long-distance travel and hypoxia adaption under high altitude environment. Meanwhile, \textit{LEF1, PRDM16} and other genes promote the occurrence of teeth and tongue, which provide physiological conditions for chewing on vegetation.
and drinking water. These characteristics are mainly related to their survival in the high-elevation environment, body consumption, warmth and foraging, etc. GO:0001654 (eye development) was mainly enriched with \textit{RBP4} and \textit{ATF6}. \textit{RBP4} and \textit{ATF6} are significantly positively correlated with rod photoreceptors, which in turn affect retinal and visual function\cite{43,44}. GO:0007162 (negative regulation of cell adhesion) was mainly enriched with \textit{MDK}. \textit{MDK} can affect the proliferation of glial cells and the formation of Muller glia-derived progenitor cells. \textit{MDK} is highly expressed in the embryonic retina and promotes the survival of retinal neurons\cite{45}. GO:0007605 (sensory perception of sound) was mainly enriched with \textit{OTOG} and \textit{OTOS}, the loss of \textit{OTOG} and \textit{OTOS} can cause hearing loss in the case of complete hearing organ structure\cite{46,47}. \textit{USH1C} is expressed in both retina and ear, but the expression level in the retina is lower than that in the ear\cite{48}. \textit{PKHD1L1} is expressed in the tip of stereocilia, especially in the high-frequency regions of the cochlea, and is necessary for normal hearing\cite{49}. GO:0042491 (auditory receptor cell differentiation) was mainly enriched with \textit{JAG1} and \textit{ATOH1}. \textit{JAG1} is expressed in the neurosensory domain and plays a role in determining the sensory and neural components of the ear\cite{50}. Reactivation of \textit{ATOH1} expression affects the number of sensory hair cells in the inner ear, allowing detection and the balancing of sounds\cite{51}. The expression of these genes can indicate that Hetian sheep in the high-altitude environment have better vision and stronger hearing than sheep raised on the plain, and thus have higher alertness. These characteristics are related to their foraging and avoiding natural enemies in the high-altitude environment.

In addition, 16 KEGG \_PATHWAY was enriched to 96 genes, and oas01100 (Metabolic pathways) were mainly enriched to 48 specific genes. Genes \textit{WNT6}, \textit{WNT10A}, \textit{LEF1} and \textit{TCF7L1} were \textit{enriched} in oas04310 (Wnt signaling pathway), oas04916 (Melanogenesis), oas05217 (Basal cell carcinoma pathway). The expression of \textit{WNT6} and \textit{WNT10A} can stimulate the formation and differentiation of osteoblasts and play an important role in maintaining bone homeostasis\cite{31}. Meanwhile, bone differentiation can promote the formation of the Generation of Osteocalcin and regulate tooth formation\cite{52}. On the other hand, WNT signaling promotes the proliferation of melanocyte stem cells and drives their differentiation to produce pigment\cite{53}. \textit{LEF1} can secrete various growth factors, which have a certain protective effect on the heart, and can also stimulate osteoblast differentiation and promote tooth development\cite{34,55}. \textit{LEF1} is also highly expressed in melanoma and melanin\cite{53,56}. \textit{TCF7L1} not only directly regulates cardiomyocyte differentiation and promotes hair follicle development, but also affects eye development\cite{57-59}. \textit{WNT6}, \textit{WNT10A}, \textit{LEF1} and \textit{TCF7L1} mainly regulate body growth, heart, teeth and wool, and provide a certain basis for the adaptation of Hetian sheep to migration, hypoxia, foraging and warmth in high altitude environments. At the same time, \textit{LEF1} promotes melanin expression can protect the deeper layers of skin and eyes from ultraviolet radiation in the high-altitude environment, and has a protective effect on Hetian sheep. However, when the growth of excessive melanocytes is out of control, cancer will occur, causing melanoma.

\textbf{Exploring the regulatory mechanisms of low-altitude economic pathways}
For further study the genetic characteristics and physiological mechanism of MGTS and GTS on grass environment. In this study, a total of 191 MGTS and GTS specific genes were screened by GO and KEGG enrichment, and were enriched to 52 pathways. These genes were not expressed on MTS. The results of biological enrichment may help to identify genes related to important economic traits, and can help us better understand the biological processes and mechanisms of natural selection or artificial selection in Hetian sheep population.

Genes enriched to GO:0045669 (positive regulation of osteoblast differentiation), GO:0000082 (G1 / S transition of mitotic cell cycle), GO:0003143 (embryonic heart tube morphogenesis), GO:0000902 (cell morphogenesis), GO:0007160 (cell-matrix adhesion), GO:0007179 (transforming growth factor β receptor signaling pathway), GO:0048333 (mesodermal cell differentiation), GO:0014850 (response to muscle activity), GO:0043149 (stress fiber assembly) were mainly responsible for osteoblastic ration, proliferation and differentiation. Cellular morphogenesis, matrix adhesion, stimulation of muscle activity and various states or activities of muscle cells in organisms. YAP1 maintains normal cerebellar morphology and coordinated movement\[^{60}\]. ITGA8 leads to the expression of actin filaments\[^{61}\]. FGF9 signal transduction mechanism regulates myogenesis\[^{62}\]. IST1 regulates cell division\[^{63}\]. ITGB6 is expressed in skeletal muscle\[^{64}\]. ACVR1 is critical in bone formation and development\[^{65}\]. CDK13 maintains embryonic stem cell self-renewal\[^{66}\]. CAPN3 plays an important role in the promotion of calcium release from skeletal muscle fibers, sarcoplasmic reticulum calcium uptake, muscle formation and muscle remodeling\[^{67}\]. GLI3 regulates the activation of muscle stem cells\[^{68}\]. In addition, INHBA and EIF2S2 regulate important biological processes related to male gonad development and reproduction. INHBA, a mesenchyme-specific gene, acts collectively with testosterone to facilitate epididymal coiling by stimulating epithelial proliferation\[^{69}\], while promoting sperm production and maintaining normal testicular morphogenesis\[^{70}\]. EIF2S2 affects germ cell proliferation and differentiation\[^{71}\]. These characteristics are relate to the limitation of production mode in agricultural areas. In the selection of agricultural production and breeding, compared with genetic character, more attention is paid to the growth and development and reproductive traits of Hetian sheep, which in turn leads to a decrease in its genetic potential, so it will have a farther genetic distance from its ancestors. IFIT3 and IFIT2 enriched in GO:0006957 (complement activation, alternative pathway) are highly expressed, as proteins of the IFIT family during the cell-intrinsic immune response to viral infection\[^{72}\]. These results provide a new understanding of genomic changes caused by artificial selection and their effects on fertility, growth and development, and immune genes.

In addition, in the KEGG _ PATHWAY, PLA2G4F, ACVR1, PRKAA2, ADCY2, ADCY5, ZFHX3, and INHBA were enriched to oas04921 (Oxytocin signaling pathway), oas04913 (Ovarian steroidogenesis), oas04550 (Signaling pathways regulating pluripotency of stem cells). These are reproduction-related genes that work in the fallopian tubes, fetal weight, uterine artery diameter and other pathways. PLA2G4F releases arachidonic acid, a precursor of prostaglandin synthesis, which plays an important role in ovulation, fertilization, implantation and parturition. Prostaglandins can either contract or relax the muscle layers of the oviduct, which in turn transports gametes and zygote\[^{73,74}\]. ACVR1 is present in granulosa cells of
mid-estrus and pre-ovulation follicles and is significantly elevated in testicular proliferation\cite{75,76}. PRKAA2 affects uterine artery diameter, fetal growth, birth weight and metabolic homeostasis\cite{77}. Estradiol significantly promoted the expression of ADCY2\cite{78}. ZFHX3 plays a critical role in the proper lactation development in the mammary gland\cite{79}. ADCY5 were robustly associated not only with fetal birth weight, but also the regulation of placental glucose transporter expression, vitamin B$_2$ uptake in the placenta and the architecture and permeability of the materno-fetal placental barrier\cite{80}. The genes enriched in oas05414 (Dilated cardiomyopathy), oas05410 (Hypertrophic cardiomyopathy), oas04611 (Platelet activation pathway): ITGB5, ITGB6, ITGA8, ADCY2, ADCY5, PRKAA2, MYLK, PLA2G4F, SNAP23, which were related to inflammatory response. For example: SNAP23 is essential for platelet and mast cell development and required in connective tissue mast cells for anaphylaxis\cite{81}. PLA2G4F is a candidate gene for inflammatory bowel disease\cite{82}. MYLK mediates a regulatory pathway in neutrophils during innate immune responses\cite{83}. ITGB5, ITGB6, ITGA8, ADCY2, ADCY5, PRKAA2, MYLK, SSTR1, CREB5, FGF9, MDM2 and GSN enriched in pathway Oas04810 (Regulation of actin cytoskeleton), oas04151 (PI3K-Akt signaling pathway), oas04935 (Growth hormone synthesis, secretion and action), oas04927 (Cortisol synthesis and secretion) can regulate cell growth, development and metabolism, actin cytoskeleton regulation, cell transcription, translation, proliferation, growth and survival in Hetian sheep. Meanwhile, it has a certain regulatory mechanism for the growth and development of the body and the growth of muscle cells. Peptide hormone SSTR1 is a potential candidate gene for livestock growth traits and regulates cell proliferation, growth hormone release and other processes\cite{84}. CREB5 plays a role in maintaining articular chondrocyte expression\cite{85}. FGF9 is essential for skeletal development and is highly expressed in mature osteoblasts\cite{86}. GSN regulates actin cytoskeleton remodeling, etc\cite{87}.

These pathways indicate the important economic traits of Hetian sheep in reproduction, growth and development and immunity after artificial selection and breeding.

**Conclusions**

Based on the Illumina Ovine SNP50 BeadChip data of Hetian sheep, the genetic structure and ROH within the population were analyzed. According to the geographical altitude difference, it was found for the first time that Hetian sheep can be divided into 3 subtypes: MTS, MGTS, GTS at the genome level. There is a closer genetic distance between GTS and MGTS. MGTS is a new strain evolved by some GTS to adapt to the living environment of MTS under the action of natural selection and artificial breeding. Therefore, it has the genetic characteristics of MTS and GTS. The current diversity distribution of Hetian sheep gene structure is due to its multiple migration and artificial selection in different historical periods. Simultaneously, ROH analysis found that there were genes related to multiple birth traits in GTS and MGTS, which provided some ideas for the breeding of multiple birth lines in Hetian sheep.

**Declarations**

Ethics approval and consent to participate
The authors declare that the experimental protocol was reviewed and approved by the College of Animal Science and Technology of Tarim University (approval number TARU-HTS-2022-001) and performed in agreement with the Care and Use Guidelines of Experimental Animals published by the Ministry of Science and Technology of the People’s Republic of China.

The authors declare that that animal tissue samples were collected in accordance with the veterinary practice regulations of the country where the samples were taken.

The authors declare that all methods are reported in accordance with ARRIVE guidelines for the reporting of animal experiments.

Data Accessibility

All data generated or analysed during this study are included in this published article and its supplementary information files.

Conflict of Interest Statement

The authors declare that the research was conducted without any potential conflict of interest.

Author Contributions

Z.-P.H. and S.-D.L. conceived and supervised the study. Z.-P.H. analyzed the data and W.Z. also contributed. L.-L.Z., R.-T.W., C.-J.L. and X.-Y.B. contributed samples or provided help during the sample collection. Z.-P.H. and S.-D.L. wrote the paper. All the authors reviewed and approved the final manuscript.

Acknowledgments

We would like to thank everyone who contributed towards the article

**Funding.** The study was funded by grants from the National Natural Science Foundation of China (NO: 32060743), Bingtuan Science and Technology Program (2022CB001-09) and Ting di joint participation in major science and technology projects.

Contributor Information

Zhipeng Han, Email: 2013355246@qq.com

Wen Zhou, Email: 1903384650@qq.com

Lulu Zhang, Email: 1463493011@qq.com

Ruotong Wang, Email: 374191227@qq.com

Chenjie Liu, Email: guilt369@163.com
References


Figures

![Map of Hetiansheep sampling points](image)

**Figure 1**

Sampling point of 84 Hetiansheep
**Figure 2**

Hetian sheep PCA diagram. 2a is PC1 and PC2 cluster analysis. 2b is PC1 and PC3 cluster analysis. 2c is PC2 and PC3 cluster analysis.

**Figure 3**
Evolutionary tree diagram of Hetiansheep. From the evolutionary tree, it can be seen that the Hetiansheep are divided into three groups: MTS, MGTS and GTS

**Figure 4**

Hetian sheep admixture diagram. K=2, MGTS shares a common ancestor with GTS. K=3, each of the three subgroups has a different ancestor. K=4, a new unknown group appears within MTS

**Figure 5**
Linkage imbalance (LD) attenuation plot of three groups of Hetiansheep in kb

**Figure 6**

Study of the correlation between length and number of long homozygous fragments

**Figure 7**
Ratio of the number of long homozygous fragments (ROH) and the longest homozygous fragments (ROH) on autosomals in each population

Figure 8

Three population candidate gene Wayne diagrams

Figure 9
Figure 10

GO enrichment scores of candidate genes associated with MTS and MGTS

Figure 11

Dilated cardiomyopathy
Platelet activation
Regulation of actin cytoskeleton
Oxytocin signaling pathway
Longevity regulating pathway
Hyper trophyic cardiomyopathy
TGF-beta signaling pathway
PI3K-Akt signaling pathway
Growth hormone synthesis, secretion and action
Longevity regulating pathway - multiple species
Apelin signaling pathway
Ovarian steroidogenesis
Vascular smooth muscle contraction
Cortisol synthesis and secretion
cAMP signaling pathway
Signaling pathways regulating pluripotency of stem cells
KEGG chord diagram of GTS and MGTS-related candidate genes

Figure 12

GO enrichment analysis of GTS and MGTS-related candidate genes

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx
- Additionalfile.xlsx