



OPEN

Analysis on the desert adaptability of indigenous sheep in the southern edge of Taklimakan Desert

Cheng-long Zhang^{1,2}, Liu Chunjie^{1,2}, Zhang Jihu^{1,2}, Zheng Langman^{1,2}, Chang Qianqian^{1,2}, Cui Zilong^{1,2} & Liu Shudong^{1,2}✉

The southern margin of the Taklimakan Desert is characterized by low rainfall, heavy sandstorms, sparse vegetation and harsh ecological environment. The indigenous sheep in this area are rich in resources, with the advantages of perennial estrus and good resistance to stress in most sheep. Exploring the molecular markers of livestock adaptability in this environment will provide the molecular basis for breeding research to cope with extreme future changes in the desert environment. In this study, we analyzed the population genetic structure and linkage imbalance of five sheep breeds with three different agricultural geographic characteristics using four complementary genomic selection signals: fixation index (FST), cross-population extended haplotype homozygosity (xp-EHH), Rsb (extended haplotype homozygosity between-populations) and iHS (integrated haplotype homozygosity score). We used Illumina Ovine SNP 50K Genotyping BeadChip Array, and gene annotation and enrichment analysis were performed on selected regions of the obtained genome. The ovary of Qira Black sheep (Follicular phase, Luteal phase, 30th day of pregnancy, 45th day of pregnancy) was collected, and the differentially expressed genes were screened by transcriptomic sequencing. Genome-wide selective sweep results and transcriptome data were combined for association analysis to obtain candidate genes associated with perennial estrus and stable reproduction. In order to verify the significance of the results, 15 resulting genes were randomly selected for fluorescence quantitative analysis. The results showed that Dolang sheep and Qira Black sheep evolved from Kazak sheep. Linkage disequilibrium analysis showed that the decay rate of sheep breeds in the Taklimakan Desert was higher than that in Yili grassland. The signals of FST, xp-EHH, Rsb and iHS detected 526, 332, 308 and 408 genes, respectively, under the threshold of 1% and 17 overlapping genes under the threshold of 5%. A total of 29 genes were detected in association analysis of whole-genome and transcriptome data. This study reveals the genetic mechanism of perennial estrus and environmental adaptability of indigenous sheep breeds in the Taklimakan Desert. It provides a theoretical basis for the conservation and exploitation of genetic resources of indigenous sheep breeds in extreme desert environment. This provides a new perspective for the quick adaptation of sheep and other mammals to extreme environments and future climate changes.

Sheep have been domesticated for about 11,000 years¹, providing humans with products such as meat, milk, skin and wool. Breed choice for sheep has been based on their production performance and reproductive capacity in grassland and plain areas, but less in a desert environment.

The Taklimakan Desert is located north of the Kunlun Mountains and has a harsh ecological environment². This area is considered unfavorable to sheep breeding due to the high salinity, strong solar radiation, large temperature difference between day and night and forage shortage. Analyzing the adaptation of sheep breeds to extreme environment is of great significance for current livestock breeding and future climate changes.

Researchers have recently used female Dorset sheep from the BGI-Huada Genomics Institute and Chinese Academy of Sciences and published the sheep genome sequence in 2014. They used a combination of next-generation sequencing (NGS) and genome mapping technologies, such that the sequencing depth was 75X, and the full length of the resulting genome was 2.61 Gb³. The advancements in sheep genome sequencing provide convenient genetic information for the functional gene location, analysis and related molecular markers in sheep⁴. Li et al.⁵

¹College of Animal Science and Technology, Tarim University, Alar 843300, China. ²Key Laboratory of Tarim Animal Husbandry Science and Technology, Xinjiang Production and Construction Corps, Alar 843300, China. ✉email: liushudong63@126.com

applied a combination of genome-wide association analysis and selective cleaning on 99 Finnsheep populations of different colors to obtain the genes related to the coat color (TYRP, ASIP, MITF). Based on Illumina Ovine SNP 50K Genotyping BeadChip, Kijas et al.⁶ used the FST method to carry out selective cleaning among 74 sheep breeds to obtain genes related to pigment deposition (KIT, MITF), body size, muscle hypertrophy, hair quality and production traits (NPR2, HMGA2, BMP2, FGF5); they proved that the ASIP gene is related to the black and white hair color of the sheep. A selective sweep analysis between two different types of sheep fat tail breeds was performed by Moradi et al.⁷ based on the Illumina Ovine SNP 50K Genotyping BeadChip, and they found that fat-tailed sheep have selection signals in OAR5 and OARX, while lean-tailed sheep have selection signals in OAR7. In the work of Z. Yuan et al.⁸, two selective cleaning methods, FST and hapFLK, were used to analyze fat-tailed and lean-tailed sheep populations, and candidate genes related to tail traits were obtained, including HOXA11, BMP2, PPP1CC, SP3, SP9, PROKR1 and ETAA1. Wang and colleagues⁹ performed a genomic selective sweep analysis of three sheep breeds (Mongolian, German Mutton Merino, African white Dorper) and obtained genes related to body index (APOBP, GTO), meat quality traits (ALDOA, STK32B, FAM190A), reproductive traits (CCNB2, BXP2 and SLC8A3), hair quality traits (EDAR) and ear size (MSRB3). The Illumina Ovine SNP 50K Genotyping BeadChip was further used by Wei et al.¹⁰ on six breeds of sheep populations to divide them into two groups according to the living environment of different altitudes using FST and xp-EHH analysis; as a result, they obtained the adaptation-related candidate genes of EPAS1, CRYAA, LONP1, NF1, DPP4, SOD1, PPARG and SOCS2. Ji Yang¹¹ divided Chinese sheep breeds into three groups: Yunnan-Kweichow, Qinghai-Tibetan and North-Eastern based on geographic location and obtained candidate genes that are adapted to different ecological environments (IFNGR2, MAPK4, NOX4, SLC2A4, PDK1, SOCS2, CALM2, CACNA2D1, KCNJ5 and COX2) based on the sheep genome resequencing technology. Wang et al.¹² performed an analysis on Tibetan sheep, Altay sheep, Duolang sheep, Hu sheep and Mongolian sheep using the method of sheep genome resequencing, and found 22.425–22.575Mb on the OAR22 chromosome to be involved in the regulation of blood hemoglobin levels. The CYP17 gene is associated with the number of red blood cells and high-altitude polycythemia, while DNAJB5 can improve the cell tolerance to stress factors (most notably heat stress) on the OAR2 chromosome in Duolang sheep^{13,14}. However, there is little research on the genetic mechanism behind the adaptability of sheep breeds in extreme desert environments.

With the progressive global warming and intensified desertification, extreme weather events occur frequently. Since the domesticated sheep breeds in temperate zone have poor environmental tolerance, some of them are endangered, and genetic resources are lost. Understanding the genetic mechanism behind the adaptation to different environments in sheep breeds can help to develop strategies to protect endangered breeds. In this work, we investigated the genetic mechanism of perennial estrus and environmental adaptability of indigenous sheep breeds in Taklimakan Desert by combining genome and transcriptome analysis of five sheep breeds with three different agro-geographical characteristics. This provides a theoretical basis for the development and protection of sheep germplasm resources in extreme desert environments.

Materials and methods

Ethics statement. All animals were handled according to the Guidelines for the Biological Studies Animal Care and Use Committee, People's Republic of China. Animal experiments were approved by the Animal Ethics Committee of the College of Animal Science and Technology of Tarim University.

Animal samples. A total of 314 healthy unrelated adult ewes belonging to five sheep breeds from three different agro-geographical zones were included. The sheep breeds were divided into three groups. The first group from Taklimakan Southern Edge of Desert, including 36 Qira Black sheep (QR) from Qira Sheep Farm, Qira County, Xinjiang and 21 Duolang sheep (DL) from Maigaiti Sheep Farm, Maigaiti County, Xinjiang. The second group from Yili Grassland, including 66 Kazakh sheep (HSK) from Gongliu sheep Farm, Gongliu County, Yili, Xinjiang. The third group from the International Sheep Genomics Consortium (ISGC, <http://www.sheephap-map.org>), including 103 Dorset sheep (APD) and 88 Suffolk sheep (ASU). RNA-seq samples were provided by Luyuan Bio-technology Co., Hotan Prefecture, Xinjiang. We selected healthy Qira black sheep (QR) as well as ovarian tissues at follicular stage (FP), luteal stage (LP), 30 days of gestation (P30) and 45 days of gestation (P45). Three tissue samples were replicated in each period of RNA-Seq and qRT-PCR, and three biological replicates were performed in each experiment.

Genotyping and quality control. We collected 0.25g tissue from the ear margin of sheep using ear pliers and placed the sample in a 1.5-mL sterilized centrifuge tube containing 75% alcohol. DNA was extracted using the phenol-chloroform method, and Illumina Ovine SNP50 BeadChip was prepared. The Genome Studio software was used to process the basic data and obtain VCF files, while PLINK software v.1.905¹⁵ was used to evaluate the quality of single nucleotide polymorphism (SNP) genotype data and remove unqualified SNPs from the included samples. We used the following criteria: inclusion detection rate less than 90%, minimum allele frequency less than 5% and Hardy-Weinberg balance test P value less than 1×10^{-6} .

Genetic diversity and population structure. Principal component analysis (PCA) was conducted on SNPs data after quality control using the software PLINK¹⁵. The VCF2Dis software (<https://github.com/BGI-shenzhen/VCF2Dis>) was used to calculate the P distance matrix, based on which a neighbor-joining phylogeny was constructed. The tree was visualized using the iTOL tool¹⁶. To perform admixture analysis, LD pruning was carried out in a sliding window of 50 SNPs with a step size of 10 SNPs by PLINK¹⁵. All SNPs with an r^2 of LD in each window exceeding the threshold of 0.1 (-indep-pairwise 50 10 0.1) were deleted, and the SNPs with the highest MAF in the PAIR of SNPs in LD were retained. After pruning, we selected the K value within the

range 2–14 to perform population genetic structure analysis using Admixture¹⁷. In order to determine the most probable ancestral populations, the lowest cross-validation error of 14-fold was applied. LD (average value of r^2) between SNPs among varieties was estimated using the PopLDdecay software v3.41¹⁸, and the results were visualized using a Perl script.

Selection sweep methods. In order to capture the genomic selection signals more accurately, more than one method is usually needed¹⁹. Therefore, we implemented four complementary statistical tests: FST (fixation index), xp-EHH (cross-population extended haplotype homozygosity), Rsb (extended haplotype homozygosity between-populations) and iHS (integrated haplotype homozygosity score). We performed Intra-population genomic selective sweep analysis using iHS on all individuals (QR and DL) in the first group, and pairwise comparison for (a) QR and DL versus HSK, and (b) QR and DL versus APD and ASU to identify genomic regions under increasing differentiation using FST, xp-EHH and Rsb.

The iHS method uses a single marker locus to replace the core haplotypes in EHH statistics, defining them as core loci. The ancestors in the core loci alleles in the haploid type are extended, and new mutant alleles in extension EHH statistics of haploid genetic distance integral are obtained, and the ratio between them is calculated to select the signal detection statistic.

FST analysis is a widely used method²⁰ to identify genetic differences between populations compared with the frequency of polymorphism within the populations. The FST value represents the basic measure of genetic difference between two populations. We performed the FST analysis to identify genomic regions that are under increasing differentiation using VCFtools v0.1.15. For each comparison, the average FST value was calculated across all 45,298 SNPs.

The xp-EHH test²¹ can be used to perform a complete selective scan, by comparing the corresponding haplotype of each population with that of other populations.

The Rsb²² test is used to identify selective sweeps based on the same idea of estimating EHH as the in the xp-EHH test. However, unlike the xp-EHH test, Rsb does not require phasing information²³. Similar to other statistics that usually focus on comparing genetic variation between populations, Rsb compares EHH of the same allele between different populations.

In this work, the iHS, xp-EHH and Rsb scores for each locus were calculated using the rehh package²⁴ in R, and the candidate genomic regions under selection were obtained.

For each selection scan test, we defined the thresholds of 1% and 5% for the results to detect selected regions in the genome, with annotations referencing the sheep genome Ovis Oar_v4.0. We used UpSetR²⁵ to analyze and visualize the cross-set relation of each gene set. Gene functional annotation was performed referencing the NCBI databases (<http://www.ncbi.nlm.nih.gov/gene>) and OMIM database (<http://www.ncbi.nlm.nih.gov/omim>).

Bioinformatics analysis. STRING (<http://string-db.org/>), Gene Ontology database (GO, <http://geneontology.org>)²⁶, Kyoto Encyclopedia of Genes and Genomes (KEGG, <https://www.kegg.jp>)²⁷, Reactome²⁸ and Human Phenotype Ontology (HP)²⁹ were used for protein interaction, biological function and pathway analysis, respectively.

RNA-Seq analysis. RNA was extracted from the tissues using the Qiagen RNeasy Plus Universal Kit. We prepared a kit sequencing library from KAPA Stranded RNA-Seq, and 100 bp double-ended reads were sequenced using an Illumina Next-Seq machine with a target of 50 million reads per sample. Bowtie2 was used to compare clean reads with the reference sequence. Then, RSEM was used to calculate the expression levels of genes and transcripts and establish the model of reads generation. Next, the maximum likelihood method was used to determine how to allocate reads to different transcripts. To achieve more accurate quantification, reads can be distinguished from positive and negative chains. In order to compare the expression levels between samples, it is necessary to standardize the gene expression levels. The standardized method used by RSEM is FPKM, and its calculation formula is as follows:

$$\text{FPKM} = \frac{10^6 C}{NL/10^3}$$

Let FPKM(A) be the expression quantity of gene A, then C represents the number of fragments uniquely compared with gene A, N denotes the total number of fragments uniquely compared with the reference gene, and L is the number of bases in the encoding region of gene A. Using the FPKM method, the influence of gene length and sequencing amount difference in the gene expression can be eliminated, and the calculated gene expression amount can be directly used to compare gene differential expression among different samples.

The analysis of gene selection sweep relation with RNA-Seq. The rows and columns of the FPKM constructed matrix were transposed to obtain a new matrix, in which each column corresponds to a sample. Then, we fitted this model through the least-squares method to calculate the t-statistic of the differential expression of each gene according to a standard method. The calculation formula is as follows:

$$t = \frac{(X^T X)^{-1} X^T Y[0]}{\sqrt{\text{MSE} \times (X^T X)^{-1}[0, 0]}}$$

where MSE is the mean square error of the fitted model:

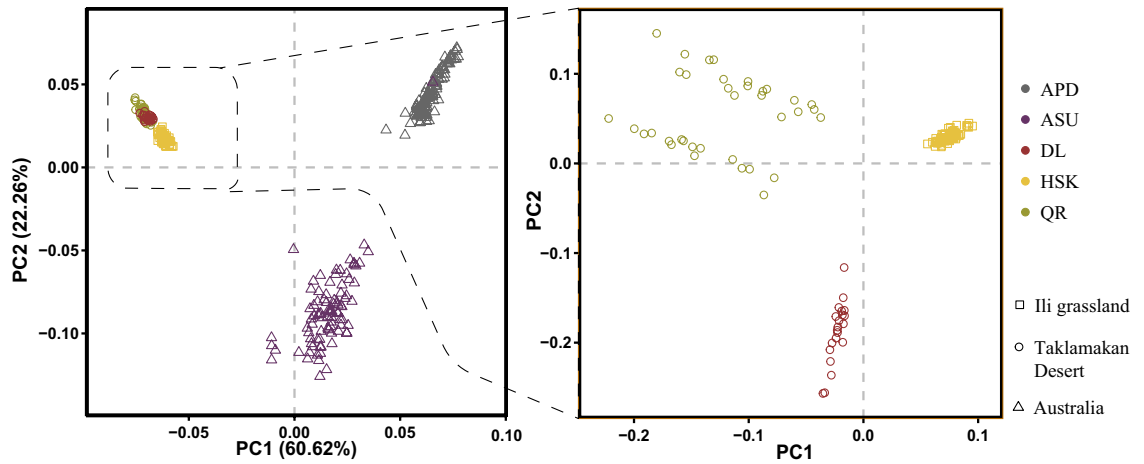


Figure 1. The PCA analysis chart of sheep breeds in three geographic regions, the X-axis represents PC1, and the Y-axis is PC2. The picture on the left shows the genetic distances of five sheep breeds, and that on the right shows the genetic distances of three sheep breeds in the Taklimakan Desert and the Yili Grassland.

$$MSE = \frac{1}{N} (Y - X(X^T X)^{-1} X^T Y)^T (Y - X(X^T X)^{-1} X^T Y)$$

where N denotes the number of rows in x . The genes were sorted according to t -statistic results. The first 7% genes of each sample were defined as the specific expression gene set (SEGS) of the sample, and the result of the selective sweep of the genome was defined as the candidate gene set (CGS). Next, we took the intersection of the candidate gene set and the specific expression gene set of each sample and clustered them based on the data analysis of the ovaries in different periods while removing redundant results.

qRT-PCR. A total of 15 genes were randomly selected from the genome selection signal results and transcriptome data analysis results. PRIMER5 was used to design quantitative primers. The primer sequences are shown in the Appendix Table S1, and the sheep 18s gene was used as an internal reference. We used the $2^{-\Delta\Delta Ct}$ analysis method to calculate the gene expression level³⁰. For each sample, three biological replicates were used. By comparing the gene expression levels in different processed samples, the ratio of each gene expression was obtained.

Results

Population genetic structure and linkage disequilibrium. After performing genome quality control, a total of 314 individuals and 45,298 SNPs were used for data analysis. As shown on the right side of Fig. 1, PCA divided the five breeds into three groups. In order to show the group classification more clearly, the Taklimakan desert environment sheep breeds and the Yili grassland sheep breeds were separately analyzed (Fig. 1right). The principal component (PC1) could separate the Duolang sheep, Qira Black sheep and Kazakh sheep.

The results of adjacent evolutionary tree are shown in Fig. 2. Five sheep breeds were well differentiated according to the geographical regions. The Taklimakan Desert sheep breed is a branch, while the Yili grassland sheep breed represents a separate branch. The evolutionary tree shows that the Duolang sheep and Qira Black sheep are more closely related. When using the outer group APD as the root of the tree, the evolutionary tree shows that the Duolang sheep and Qira Black sheep evolved from the Kazakh sheep.

In the admixture analysis, 28,568 of the 45,140 SNPs were deleted, and 16,572 SNPs were retained after LD pruning. The lowest CV error was detected when $K = 8$. QR, DL and HSK began to separate at $K = 6$, while APD and ASU began to separate at $K = 3$. When $K = 5$, QR, DL and HSK shared two main components. With respect to the size of the components, the evolution direction of the three was as follows: HSK→DL→QR. As K increased, QR, DL and HSK no longer produced obvious separation. The separation was mainly manifested in the two varieties of APD and ASU, indicating that APD and ASU may represent two artificially domesticated hybrids (Fig. 3).

The average LD coefficient r^2 of the three flock breeds QR, DL and HSK rapidly decayed within 10 kb. However, compared with the Taklimakan Desert sheep breeds, the Kazakh sheep from the northern foot of the Tianshan Mountains showed a faster decay. There was also a big difference with DL in the range of 50–300 kb (Fig. 4).

Selection sweep, gene annotation and functional analysis. The intersection and union relations of all selective scanning results of the iHS, FST, Rsb and xp-EHH methods are shown in Appendix Table S2. iHS was used to detect the selected areas of QR and DL (The results of QR are shown in Appendix Table S3 and those of DL are shown in Appendix Table S4, and their Manhattan diagrams are shown in Fig. 5). Common selected areas of QR and DL were mainly concentrated in OAR1, OAR3, OAR6, OAR9, OAR10, OAR19 and OAR26 at a threshold of 1%. Among the iHS screening genes, there were 526 intersection genes between QR and DL. These genes are related to the growth and development traits of organisms. For example, ACVR22, FADS2, FAR2, RARB, COL3A1, RERE, ITGAV, TEK and ADAMTS19 are related to animal cell development

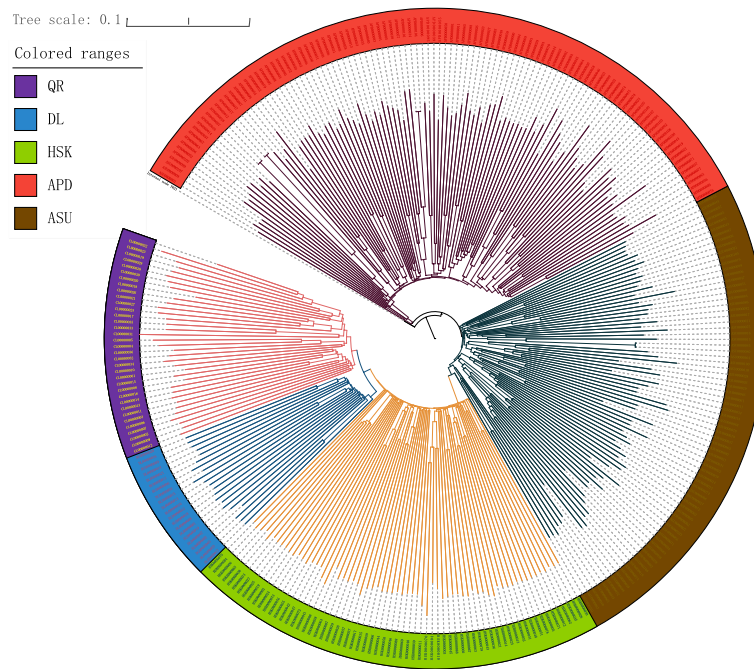


Figure 2. Evolutionary tree of five sheep breeds. Red represents APD, brown represents ASU, light green represents HSK, blue represents the DL and purple represents QR.



Figure 3. Admixture analysis of five sheep breeds. From left to right are DL, QR, HSK, ASU and APD.

and metabolism^{31–33}; KCNN3, FSHR, IGFBP7, CDC25B, CREB and FMR1 are related to early maternal follicular development and ovarian growth^{34,35}; STX, KLHL1, BIN1, SCOT and KCNMB3 affect the ion channel function^{36,37}. In the FST, Rsb, and xp-EHH tests, we used two paired comparisons: (a) QR and DL with HSK, (b) QR and DL with APD and ASU, and the results are shown in Fig. 6 and Fig. 7, respectively.

In the FST test, the following was obtained at a threshold of 1%: (a) 277 genes were detected in QR, DL and HSK, (b) 312 genes were detected in QR, DL, APD and ASU. The results of the FST test are shown in Appendix Tables S5 and S6. There were 63 gene intersections between (a) and (b), among which IHH, SORCS3, KCNMA1,

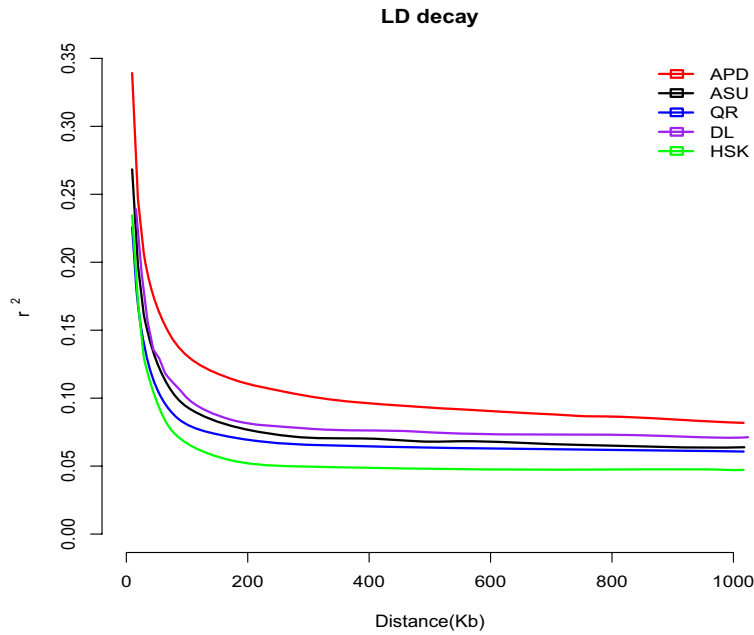


Figure 4. LD attenuation diagram of 5 sheep breeds. The X-axis represents the physical distance, while the Y-axis represents the linkage disequilibrium coefficient. Red represents APD, black represents ASU, blue represents QR, purple represents the DL, and cyan represents HSK.

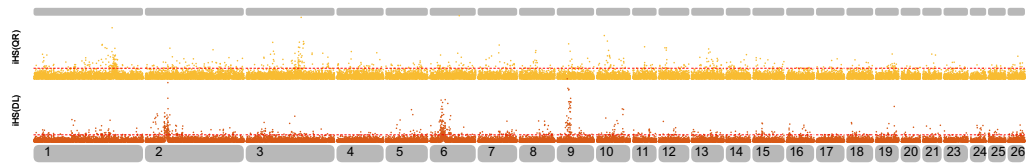


Figure 5. iHS Manhattan chart of QR and DL sheep breeds. The X-axis represents the chromosome, and the Y-axis represents the iHS value. The yellow part represents QR, and red represents DL. The red dashed line is the 1% threshold line.

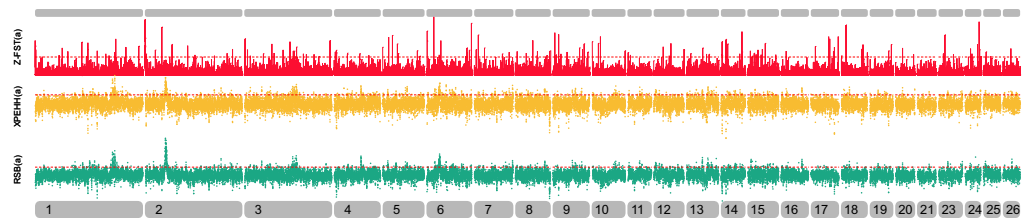


Figure 6. Manhattan chart of QR and DL versus HSK selective dissection analysis. The x-axis represents the chromosome, the green is the F_{ST} value standardized by the z-score, the orange is the xp-EHH value, and the black is the Rsb value.

TMTC1, EXT2 and HMGA2 are related to mammalian growth^{38,39}; DSG4, RORA and LDLR are related to wool quality⁴⁰; RNF157, CDH12, SIX2, SLC1A7, RXFP2 and R3HDM1 are related to adaptability in extreme environments^{41–43}.

In the Rsb test, the following was obtained under the 1% threshold: (a) QR, DL and HSK detected 171 genes, (b) QR, DL, APD and ASU detected 175 genes. The Rsb test results are shown in Appendix Tables S7 and S8. There were 37 gene intersections between (a) and (b), among which CACNB2, CLDN23, CDC226, COMMD10, EIF2B5, NDUFAF2, OR10X1 and TENM2 are related to biological adaptability in extreme environments^{44,45};

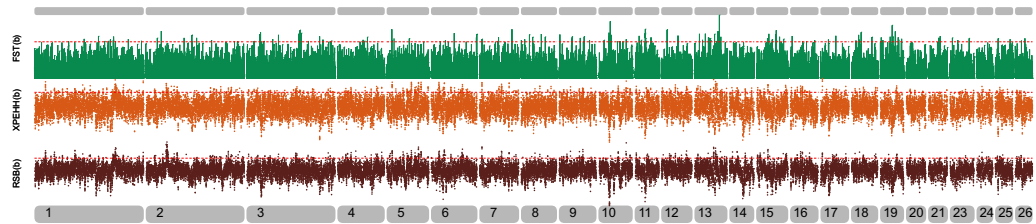


Figure 7. Manhattan chart of QR and DL vs. APD and ASU selective dissection analysis. The x-axis represents the chromosome, the green is the FST value standardized by the z-score, the orange is the xp-EHH value, and the black is the Rsb value.

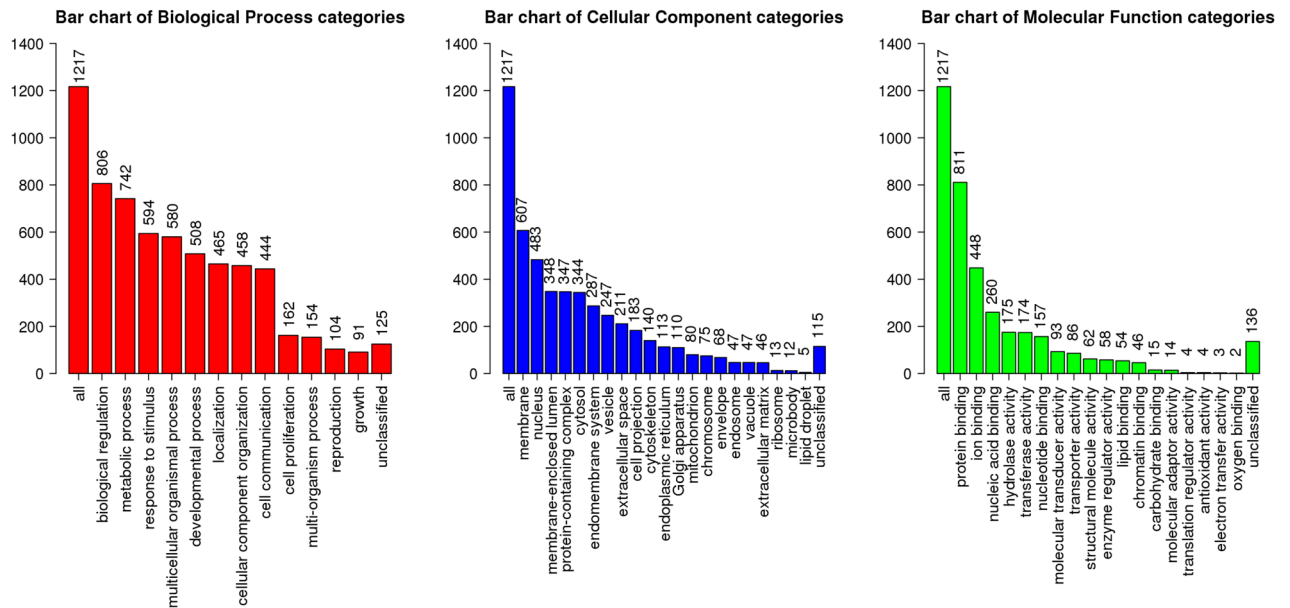


Figure 8. Selective sweep gene GO analysis at the top 1% threshold, from left to right are biological processes, cell composition and molecular functions.

CLCN2 and SLC19A1 are related to ion transport⁴⁶; ELOVL7, EP400 and MAPK8IP1 are involved in the sheep growth and development^{47–49}.

In the xp-EHH test, the following was obtained under the 1% threshold: (a) QR and DL versus HSK detected 178 genes, (b) QR and DL versus APD and ASU detected 179 genes. The xp-EHH test results are shown in Appendix Table S9 and S10. There were 25 gene intersections between (a) and (b), among which KHDRBS2, LRP4, NEIL2 and SUV39H2 are related to biological adaptability in extreme environments^{50,51}. Under the 5% threshold, 17 overlapping candidate genes were detected in IHS, FST, Rsb and xp-EHH tests, including ACVR2A, ATP11B, CCSER1, FXR1, GRID2, IGFBP7, KCNMB2, LRRC4C, MAT2B, NOX3, PAX1, RPL3, SPATA31E1, STPG2, TMT2C, USP25 and VRK2. GO enrichment analysis of significant biological processes for candidate genes under positive selective pressure revealed 51 Gene Ontology (GO) terms, shown in Appendix Table S11 and Fig. 8; the KEGG analysis resulted in 6 pathways, shown in Appendix Table S12. Furthermore, we obtained 50 Reactome pathways, listed in Appendix Table S13, and 2 Human Phenotype Ontology (HP) terms, as shown in Appendix Table S14.

The analysis of gene selection sweep relation with RNA-Seq. We conducted correlation analysis between the results of RNA-Seq (shown in Appendix Table S15) and selection genome sweeping. As a result, we determined the top 7% of each sample's specific expression gene set and its contribution (Fig. 9, Appendix Table S16). In the association analysis, the specific expression gene set at the follicular stage and the candidate gene set were found to share 17 intersection genes, of which NRG4 specifically binds to the EGF receptor ErbB4 (v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4). It exerts its physiological functions, such as stimulating cell proliferation, inhibiting cell apoptosis and improving the cell energy metabolism. For example, these functions include delaying female puberty⁵² and adult reproductive ability⁵³. In this case, the overexpression of B4GALNT2 in the ovary leads to atypical glycosylation of inhibin (inhibin is an important hormone that regulates ovarian function), which promotes high fecundity in sheep⁵⁴. In addition, previous studies have shown that this gene is related to follicular development and atresia⁵⁵. EEPD1 maintains the genome stability during

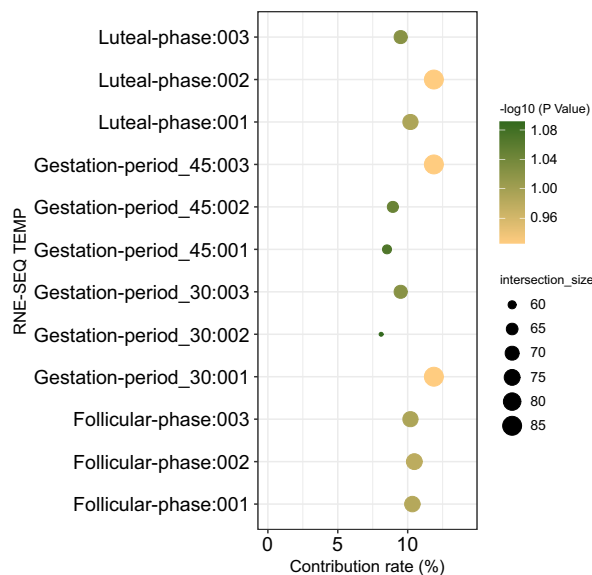


Figure 9. Candidate genes combined with selective sweeping and RNA-Seq analysis. The size of the scattered dots represents the number of genes, while the color represents the contribution of the specific gene set in each period to the total gene set.

embryonic replication stress by promoting homologous recombination (HR) and inhibiting non-homologous end joining (NHEJ)⁵⁶. The INTS9 complex promotes the recruitment of cytoplasmic dynein to NE and participates in RNA processing⁵⁷. DNMT2 plays a role in mediating endocytosis events that are critical to oocyte development; thus, it affects fertility⁵⁸, MCAM is involved in melanin Invasion and tumor progression⁵⁹ and mediates cell-endothelial cell interaction in the ovary⁶⁰; this molecule also plays a role in the formation of new blood vessels during the formation of the corpus luteum in the human ovary⁶¹. As for OPN5, it belongs to an independent group separated from the other six groups in the phylogenetic tree of opsins, for which little information of absorption characteristics and molecular properties of the members is known⁶². PCDH17 can affect the sex differentiation of the gonadal through its expression level⁶³. There were three shared genes between the specific expression genes in the luteal phase and the candidate genes set. The shared genes include FGF3, which is involved in the regulation of the corpus luteum (CL) growth and degradation⁶⁴, CHKA, which encodes an enzyme involved in the synthesis of phosphatidylcholine⁶⁵; phospholipid Acylcholine is the precursor of many physiologically important membrane molecules, including arachidonic acid and docosahexaenoic acid. These molecules are essential for male and female fertility⁶⁶. The third shared gene is UQCRH, which is responsible for the electron transfer between cytochrome C and cytochrome C1, representing a part of the most critical step in the electron transfer cascade and being involved in the oxidative phosphorylation of mitochondria. In addition, UQCRH can protect the ovaries of animals from radiation-caused damage⁶⁷.

The intersection set between the 30 days gestation specific gene set and the candidate gene set included 4 genes. Among them, F13A1 is related to the heat stress tolerance of pregnant ewes⁶⁸, ARNTL indirectly affects the fertility of animals by affecting the circadian rhythm of the animal's ability^{69,70}. DNMBP has a role in membrane transport between the cell surface and the Golgi apparatus (through genetic similarity), and GLIS3 affects the development of organs, such as the fetal pancreas or thyroid^{71–73}.

A total of 5 genes resulted in the intersection set between the 45 days gestation specific expression gene set and the candidate gene set. Among them, HTR1A promotes the fetal-maternal tolerance of the endometrium by mediating immunosuppression⁷⁴, AS3MT (Arsenic (+3 oxidation state) methyltransferase) is a gene that depends on the sex of the fetus and represents a key enzyme in the metabolism of inorganic arsenic (iAs)⁷⁵, GNAQ is a candidate mediator of melanopsin signal transduction, which can affect the fetal hyaline and retinal vascular development⁷⁶, CCDC178 may lead to fetal atrial septal defect and cardiovascular development abnormalities if in abnormal state⁷⁷, and CNGA1 participates in the coding loop cyclic nucleotide-gated channel (CNG channel), affecting pregnancy in animals⁷⁸.

qRT-PCR. A total of 15 genes were randomly selected for qRT-PCR verification to verify their expression in ovarian tissues at three different periods. The results are shown in Fig. 10. It could be seen that ACVR2A was up-regulated during the follicular phase (FP) and down-regulated during the luteal phase (LP) and pregnancy (P30). The difference between FP versus LP and P30 were highly significant. BMPER and MRPL22 were up-regulated during FP and down-regulated during the LP and P30. The difference between FP versus LP was significant, and that between FP versus P30 was very significant. IGFBP7 was up-regulated in FP and down-regulated in the LP. The difference between LP versus FP was highly significant. OPN5 was up-regulated dur-

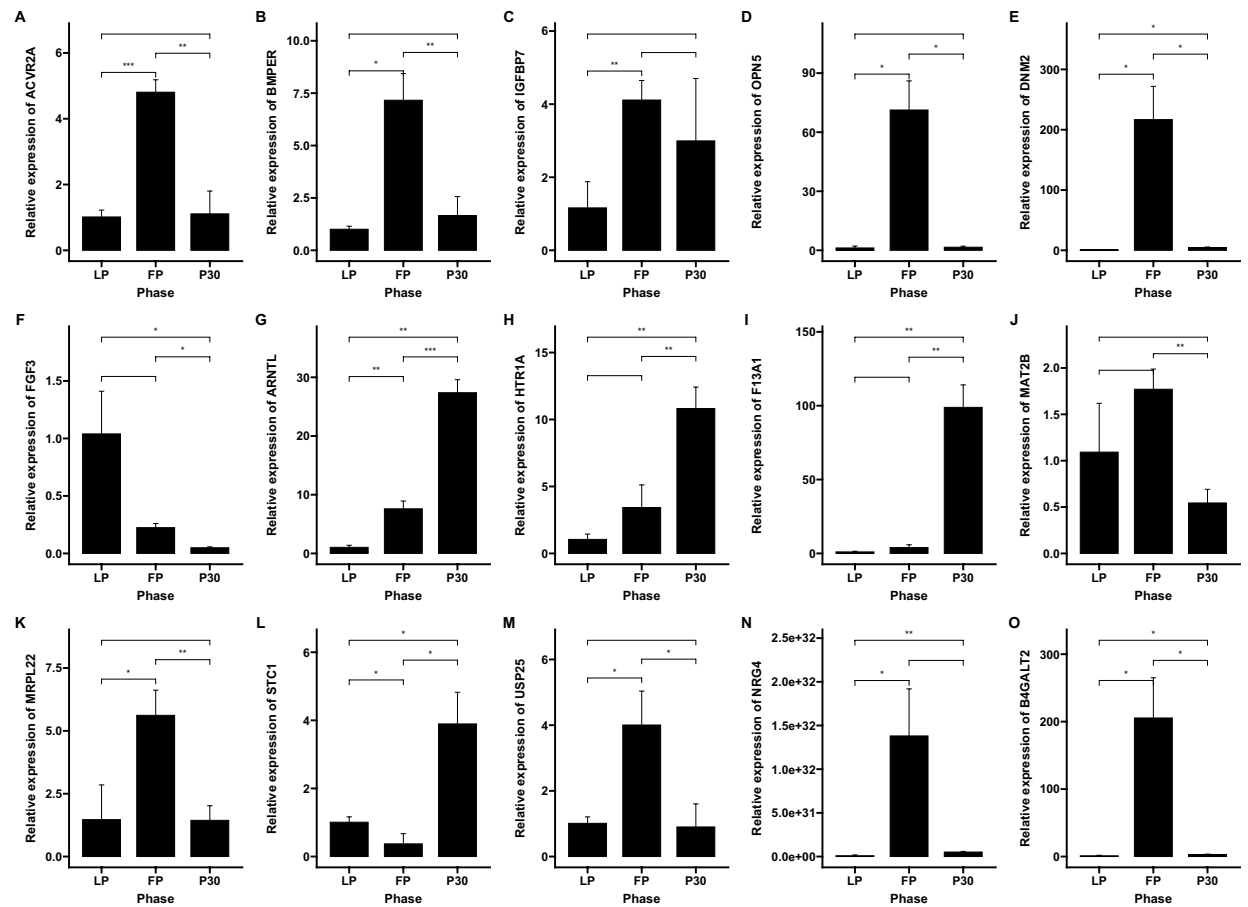


Figure 10. qRT-PCR analysis of 15 genes in the ovarian follicular phase, luteal phase and gestation period of Qira black sheep. The x-axis represents three different periods of organization, while the y-axis represents the level of gene expression. The candidate genes were selectively cleaned to detect the expression of sheep ovaries in three different periods.

ing FP and down-regulated during LP and P30. There was a significant difference between LP versus FP and a significant difference between LP versus P30. B4GALT2 and DNM2 were up-regulated during FP and down-regulated during LP and P30. There was a significant difference between LP and FP as well as between LP and P30, and a significant difference between FP and P30. NRG4 was up-regulated during FP and down-regulated during LP and P30. The difference between LP versus FP was significant, and that between LP versus P30 was very significant. FGF3 was up-regulated during LP and down-regulated during FP and P30. There was a significant difference between FP versus LP and P30, and also between LP versus P30. USP25 was up-regulated during FP and down-regulated during LP and P30. FP versus LP and P30 were significantly different. ARNTL, HTR1A and F13A1 were up-regulated during P30 and down-regulated during LP and FP. P30 versus FP and LP were significantly different. STC1 was up-regulated during P30, and down-regulated during LP and FP. P30 versus FP and LP were significantly different. MAT2B was up-regulated during FP and down-regulated during P30. The difference between FP versus P30 was extremely significant.

Discussion

Genetic diversity and population structure. PCA analysis showed the selection of sheep breeds in the Taklimakan Desert to be different from other agro-climatic breeds. The resulting cluster relationship of the five sheep breeds was closely related to their geographical distribution. The further the geographical relationship, the higher the discrimination through PCA. The evolutionary tree revealed DL and QR to be evolved from HSK. These findings are consistent with the study of Yang et al.¹¹ Admixture analysis revealed the degree of genomic mixing of the 5 sheep breeds. The ancestral components of $K = 2$ to $K = 4$ were related to the geographical origin. At $K = 2$, a small amount of mixing was detected between Australian and Chinese breeds, which might be because part of the living environment of Australia and Xinjiang sheep in China is the same. For $K = 5$, the three Xinjiang varieties were roughly divided into two parts (DL and HSK versus QR), and they were not mixed with Australian varieties. This indicates that HSK is more closely related to DL and farther from QR. When $K = 8$, the three varieties of Xinjiang could be separated. When there was still a small amount of mixing, it was mainly manifested in QR and DL. This shows that QR and DL co-evolved, and there was also a phenomenon of gene

exchange, but QR has evolved further. These results are consistent with the conclusions of PCA and adjacent evolutionary trees.

Selection sweep methods. Different tools of biological analysis have different results, but no one method can completely detect all the biological results in a genome-wide selective sweep analysis¹⁹. We improved the accuracy of our data using a variety of genomic selection methods. Selected sites of DL and QR varieties were scanned by iHS. The methods of FST, xp-EHH and Rsb were used to scan the selected difference sites of (a) QR and DL versus HSK, and (b) QR and DL versus APD and ASU. The results revealed genetic evidence of the adaptation of sheep breeds in the Taklimakan Desert to local environments. In the results of genome selective cleaning, 1,279 genes were collected under the threshold of 1% (Appendix Table S17), while there are 17 overlapping genes under the threshold of 5%.

Adaptive mechanism of the desert environment. The environment adaptability differs between the Taklimakan Desert sheep breeds and those of other areas. Taklimakan Desert indigenous sheep breeds can adapt to extreme conditions, such as high salinity, drought and high ultraviolet rays. Using four complementary genomic selective sweeps, 1279 genes were obtained under the 1% threshold, and the biological enrichment results of these genes showed the genetic evidence and physiological mechanism of the adaptation of native sheep to the desert environment.

In terms of adaptability, Ca^{2+} activated K^+ channels (R-HSA-1296052) promote cytoplasmic Ca^{2+} accumulation. Then, they reduce K^+ elimination and cytoplasmic Na^+ accumulation²⁸. cGMP affects (R-HSA-418457) by affecting phosphodiesterase (PDE) and cGMP-dependent protein kinase (cGK, Protein Kinase G or PKG); it can regulate vasodilation, platelet aggregation and neurotransmission, as physiological functions²⁸. In addition, among the target genes of the association analysis with the transcriptome, F13A1 in the 30 days gestation component is related to the heat stress tolerance of pregnant ewes⁶⁸.

In terms of immunity, 5 genes (DNMT1, FEN1, POLL, PRKDC and XRCC4) participate in the non-homologous end-joining pathway. XRCC4 and DNMT1 are involved in repairing the DSB repair caused by exogenous or endogenous DNA damage as well as the DSB repair produced in the process of V(D)J recombination⁷⁹. Eight genes (ABCC1, ABCC2, ABCD4, ABCC5, ABCC6, ABCC11, ABCC12 and ABCB10) are involved in the ABC transporters pathway. These proteins transport a variety of substrates across the membrane, and they can cause cystic fibrosis, neurological diseases, retinal degeneration, cholesterol and bile transport defects, anemia and drug reactions when mutated⁸⁰. MAPK (R-HSA-5683057) and PI3K/Akt (hsa04151) activate the macrophages induced by ASKP-1, thereby enhancing immune regulation²⁸. Under a high radiation environment, the SUMOylation of DNA damage response and repair proteins (R-HSA-3108214) pathway is involved in DNA damage response and repair²⁸.

The results of the biological enrichment of candidate genes using the Human Phenotype Ontology database were mainly related to vision. The 10 genes dominate eye development, pigment changes and light perception. Among them, MAFB regulates the lens development⁸¹. As for PTEN, it affects the cell size and inhibits the cell cycle process of early mitosis, through inactivation and overexpression, thereby promoting eye growth and development⁸². MITF plays a central role in the regulation of neuroepithelial domains and also the differentiation of retinal pigment epithelium, promoting the development of the normal retina and pigment epithelium⁸³. Regarding EDN3, it plays a role in the development of ocular angiogenesis in the early postnatal period⁸⁴. Rapid and accurate eye development represents the basis of vision development. These genes regulate eye development at different stages and positions. They are candidate genes for vision-related formation in indigenous sheep after long-term breeding. PNPLA6 encodes a protein that plays an important role in photoreceptor survival⁸⁵. SNAI2 can regulate the extracellular matrix (ECM) protein of trabecular meshwork (HTM) cells, which in turn play a key role in aqueous humor circulation⁸⁶. These genes promote the eye development of indigenous sheep and the development of their observation ability and agility. Eye pigments are produced by pigment cells, and pigments are directly related to vision and radiation resistance. The Taklimakan Desert sheep breed has long lived in a high ultraviolet environment, and the eyes and other organs have correspondingly adapted. For example, sun exposure may cause the development of melanoma, and GNAQ mutations can affect uveal melanoma (UM)^{87,88}. Some genes determine the eye color and protect melanocytes from radiation, such as ACTB, KITK and ITLG^{89,90}. In the genome-wide and transcriptome association analysis, the follicular phase-specific and significantly expressed gene OPN5 (neurospine) is a bistable pigment that is sensitive to ultraviolet light⁶², which is related to the adaptability of the Taklima Desert breeds to the high ultraviolet environment. Compared with other domesticated sheep breeds, indigenous sheep breed of the Taklimakan Desert have better vision and more alertness, and these characteristics are related their foraging and avoiding of natural enemies in extreme desertification environments.

Genetic mechanism of perennial estrus and reproduction. The animal estrus cycle is the alternate process of FP and the LP, which affects follicular development. The formation and degeneration of the corpus luteum results from the joint influence of the genes and environment. Sheep breeds underwent natural and artificial selection in the environment of the Taklimakan Desert for a long time, and have developed stable reproduction and perennial estrus.

By combining whole-genome and RNA-Seq, a series of genes that are related to development and reproduction were found, mainly in the following two aspects.

FGF3 regulates the growth and degeneration of the corpus luteum⁶⁴, PCDH17 is involved in the stabilization of sex differentiation of the gonad⁶³, and MCAM mediates the cell-endothelial cell interaction in the ovary⁵⁹. ARNTL^{69,70}, CHKA⁶⁵ and NRG4⁵³ guarantee the balance and supply of hormones and reproductive factors during sheep reproduction, thus improving the maternal reproductive ability. B4GALNT2⁵⁴ and DNM2⁵⁸ affect the

follicular development and related physiological processes, hence promoting high animal production. Indigenous sheep use these genes to increase the reproductive rate and achieve multiple birth rates, directly or indirectly and in different ways, which is necessary for the continuation of the population in harsh environments. F13A1 improves heat stress tolerance of ewes⁶⁸, while UQCRH protects the organisms from radiation-induced ovarian damage⁶⁷, which is essential for sheep living in desert environments. As for CCDC178 and CNGA1, they are involved in immune repair^{77,78}, while GLIS3^{71–73} and GNAQ⁷⁶ are involved in stable fetal organ development. EEPD1⁵⁶ plays a role in maintaining the stability of embryonic genome replication. HTR1A⁷⁴ mediates immunosuppression to improve fetal-maternal tolerance, which is essential to reduce the miscarriage rate. Indigenous sheep have evolved a more unique and stable reproductive system to ensure the stability of pregnancy in the harsh desert environment.

The Reactome results of the enrichment analysis included the TFAP2 (AP-2) family, which regulates the transcription of growth factors and their receptors (R-HSA-8866910) pathway to ensure stable cell differentiation and proliferation⁹¹. The quantitative analysis of randomly selected genes showed the STC1 gene to be highly expressed in the ovary of sheep, such that the expression level is higher in the follicular phase, significantly decreases after entering the luteal phase, and reaches the highest level at 30 days of pregnancy. STC1 is a glycoprotein hormone that appears to have autocrine/paracrine effects in several mammalian tissues⁹². It is involved in a variety of biological processes, such as apoptosis^{93,94}, inflammation⁹⁵ and reproduction⁹⁶. Previous studies have shown STC1 to be highly expressed in the ovaries of humans and mice^{97,98}. Harminder found the STC1 protein to mainly exist in the egg membrane and oocytes. Although it is expressed in the corpus luteum, the expression level is low and regulated by the luteinizing hormone (LH)⁹².

STC1 in the human follicular fluid can play a certain regulatory role when it is non-covalently bound to the pregnancy-associated plasma protein (PAPP-A), and it can inhibit the biosynthesis of progesterone downstream of adenylate cyclase⁹⁹. Treatment of granulosa cells with STC1 can reduce the stimulation of gonadotropins for progesterone production and LH receptor formation, without affecting the secretion of estradiol by gonadotropins¹⁰⁰; thus further indicates the potential role of STC1 as a luteinizing inhibitor. STC1 co-treatment significantly inhibits the transcription of the rate-limiting enzyme CYP11A of progesterone biosynthesis, induced by FSH^{101–103}. During pregnancy, STC1 levels continue to rise, and on the 10th day of pregnancy, it reaches 15 times the level of normal cycle mice, and STC1 can then be detected in the serum⁹². These reports on STC1 in the estrus cycle and pregnancy indicate that STC1 plays an important role in mammalian ovarian physiology and reproduction. This raises the need for more studies on its specific functions.

IGFBP7 is abundantly expressed in the ovaries, and its expression significantly rises after entering LP from FP and decreases after entering the pregnancy period. This gene is involved in the formation of vascular endothelial growth factor (VEGFA) and LH. VEGFA and LH stimulate the formation and proliferation of LEC (microvascular endothelial cells), which indicates that the secretion of IGFBP7 in the corpus luteum may inhibit VEGFA, thus affecting early CL angiogenesis¹⁰⁴. This gene is expressed at a high level throughout the growth of antral follicles of buffalo and may thus have an impact on fertility by regulating the growth of follicles¹⁰⁵.

Finally, the MRPL22 gene is abundantly expressed in LP. Its expression significantly increases after entering the luteal phase from the ovarian stage and significantly decreases after entering the pregnancy period. It was revealed by related research on the in-depth characterization of the mouse embryo mutant MRPL22 that the absence of MRPL22 led to severe developmental delay and the inability to initiate gastrulation on the 7.5th day of the embryo. Besides, it was shown by related metabolic analysis that the mitochondrial function of MRPL22 knockout embryos was impaired, which resulted in a significant decrease in ATP production¹⁰⁶.

In a word, the stable production of sheep requires accurate and rapid development, precise immune control and tolerance to the environment. To achieve high sheep production, follicular quality and high ovulation rate are essential. In the Taklimakan Desert, native sheep breeds have formed a unique reproduction and immune control mechanism through natural and artificial selection. This physiological mechanism allows them to survive in harsh environments. Their adaptation to the desert environment is a complex and continuous process. Figure 11 briefly describes its physiological mechanism and relationship from the aspects of reproduction, drought, salt and alkali tolerance as well as vision development.

New Views on breeding of tolerance sheep breeds. The main focus in sheep breeding is on economic benefits, e.g., cultivating new varieties with high-quality meat and wool. Genetic drift and selection represent the main two sources of genetic evolution. Random genetic drift will cause losing the genetic diversity and irreversible consequences, such that the reduced genetic diversity will limit the ability of the introduced population to evolve in its new or changing environment. To avoid this, we propose two solutions: First, adopting supportive breeding, which means bringing some individuals of wild populations into captivity for breeding, and performing artificial selection in the offspring. While this program aims to improve the environmental adaptability of domesticated populations, it has become more difficult to be implemented with the shrinking of the wild sheep population. Second, cultivating conservation breeding for economic crops. This strategy protects the species diversity in extreme environments and preserves and domesticates populations in extreme environments. These proposed measures can enhance the adaptability of species in the new environment, such that they can deal with unpredictable climate changes in the future.

Conclusions

This work used four complementary genomic selection signal analyses to reveal the genetic mechanism of stress resistance of indigenous sheep breeds in the Taklimakan Desert environment. We successfully constructed selection signal maps of sheep breeds in three different geographical and ecological environments. In addition, we elaborated on the physiological system of sheep breeds in the Taklimakan Desert, including immunity, vision

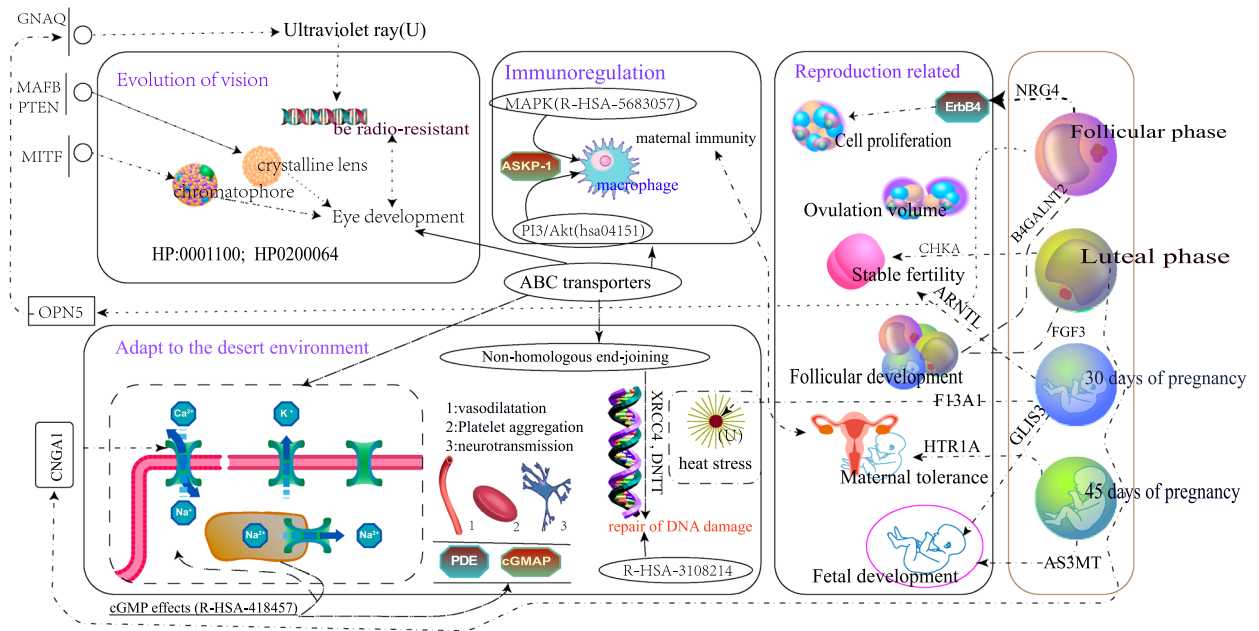


Figure 11. Environmental adaptability and perennial estrus regulatory network of the Taklimakan Desert sheep breeds.

degeneration, high reproduction rate and water reabsorption, and we constructed the prediction mechanism map of desert adaptability. Through the combined analysis of genomic selection signal and transcriptome, we analyzed the genetic mechanism of perennial estrus and stable reproduction of indigenous sheep in the Taklimakan Desert environment, and the related genes of B4GALNT2, FGF3 and F13A1 were mainly involved. These results help to understand the genetic mechanism behind the adaptability of sheep breeds in desert environment at the molecular level and provide new ideas for breeding tolerant sheep breeds in extreme environment.

Received: 29 November 2021; Accepted: 4 July 2022

Published online: 18 July 2022

References

- Zeder, M. A. Domestication and early agriculture in the mediterranean basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci.* **105**, 11597–11604. <https://doi.org/10.1073/pnas.0801317105> (2008).
- Piao, S. *et al.* The impacts of climate change on water resources and agriculture in China. *Nature* **467**, 43–51. <https://doi.org/10.1038/nature09364> (2010).
- Jiang, Y. *et al.* The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**, 1168–1173. <https://doi.org/10.1126/science.1252806> (2014).
- Luo, W. *et al.* Genome-wide association analysis of meat quality traits in a porcine large white × minzhu intercross population. *Int. J. Biol. Sci.* **8**, 580–595. <https://doi.org/10.7150/ijbs.3614> (2012).
- Li, M.-H., Tiirikka, T. & Kantanen, J. A genome-wide scan study identifies a single nucleotide substitution in ASIP associated with white versus non-white coat-colour variation in sheep (*ovis aries*). *Heredity* **112**, 122–131. <https://doi.org/10.1038/hdy.2013.83> (2013).
- Kijas, J. W. *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* **10**, e1001258. <https://doi.org/10.1371/journal.pbio.1001258> (2012).
- Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrababak, M., Dodds, K. G. & McEwan, J. C. 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. <https://doi.org/10.1186/1471-2156-13-10> (2012).
- Yuan, Z. *et al.* Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Anim. Genet.* **48**, 55–66. <https://doi.org/10.1111/age.12477> (2016).
- Wang, H. *et al.* Genome-wide specific selection in three domestic sheep breeds. *PLoS ONE* **10**, e0128688. <https://doi.org/10.1371/journal.pone.0128688> (2015).
- Wei, C. *et al.* Genome-wide analysis reveals adaptation to high altitudes in tibetan sheep. *Sci. Reports.* <https://doi.org/10.1038/srep26770> (2016).
- Yang, J. *et al.* Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.* **33**, 2576–2592. <https://doi.org/10.1093/molbev/msw129> (2016).
- Wang, W. *et al.* Deep genome resequencing reveals artificial and natural selection for visual deterioration, plateau adaptability and high prolificacy in chinese domestic sheep. *Front. Genetics* <https://doi.org/10.3389/fgene.2019.00300> (2019).
- Vjestica, A., Zhang, D., Liu, J. & Oliferenko, S. Hsp70-hsp40 chaperone complex functions in controlling polarized growth by repressing hsf1-driven heat stress-associated transcription. *PLoS Genet.* **9**, e1003886. <https://doi.org/10.1371/journal.pgen.1003886> (2013).
- Foster, N. L., Lukowiak, K. & Henry, T. B. Time-related expression profiles for heat shock protein gene transcripts (HSP40, HSP70) in the central nervous system of *Lymnaea stagnalis* exposed to thermal stress. *Commun. Integr. Biol.* **8**, e1040954. <https://doi.org/10.1080/19420889.2015.1040954> (2015).

15. Purcell, S. *et al.* Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575. <https://doi.org/10.1086/519795> (2007).
16. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245. <https://doi.org/10.1093/nar/gkw290> (2016).
17. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093. <https://doi.org/10.1534/genetics.112.145037> (2012).
18. Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M. & Yang, T.-L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788. <https://doi.org/10.1093/bioinformatics/bty875> (2018).
19. Hohenlohe, P. A., Phillips, P. C. & Cresko, W. A. Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *Int. J. Plant Sci.* **171**, 1059–1071. <https://doi.org/10.1086/656306> (2010).
20. Weir, B. S. & Cockerham, C. C. Estimating f-statistics for the analysis of population structure. *Evolution* **38**, 1358. <https://doi.org/10.2307/2408641> (1984).
21. Tang, K., Thornton, K. R. & Stoneking, M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* **5**, e171. <https://doi.org/10.1371/journal.pbio.0050171> (2007).
22. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312**, 1614–1620 (2006).
23. Weigand, H. & Leese, F. Detecting signatures of positive selection in non-model species using genomic data. *Zool. J. Linn. Soc.* **184**, 528–583. <https://doi.org/10.1093/zoolinnean/zly007> (2018).
24. Gautier, M. & Vitalis, R. rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177. <https://doi.org/10.1093/bioinformatics/bts115> (2012).
25. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364> (2017).
26. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
27. Kanehisa, M. & Goto, S. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
28. Joshi-Tope, G. *et al.* Reactome: A knowledgebase of biological pathways. *Nucleic Acids Res.* **33**, D428–D432 (2005).
29. Robinson, P. N. *et al.* The human phenotype ontology: A tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.* **83**, 610–615. <https://doi.org/10.1016/j.ajhg.2008.09.017> (2008).
30. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* **25**, 402–408. <https://doi.org/10.1006/meth.2001.1262> (2001).
31. Nicoletta, *et al.* Glut10 deficiency leads to oxidative stress and non-canonical $\alpha v \beta 3$ integrin-mediated *tgf \beta* signalling associated with extracellular matrix disarray in arterial tortuosity syndrome skin fibroblasts. *Hum. Mol. Genet.* **24**, 6769–6787. <https://doi.org/10.1093/hmg/ddv382> (2015).
32. Guti Árrrez-Gil, B., Arranz, J. J. & Wiener, P. An interpretive review of selective sweep studies in bos taurus cattle populations: Identification of unique and shared selection signals across breeds. *Frontiers in Genetics* <https://doi.org/10.3389/fgene.2015.00167> (2015).
33. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15**, 786–801 (2014).
34. Bonnet, A. *et al.* An overview of gene expression dynamics during early ovarian folliculogenesis: Specificity of follicular compartments and bi-directional dialog. *BMC Genom.* **14**, 904. <https://doi.org/10.1186/1471-2164-14-904> (2013).
35. Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight dna repair and immune pathways. *Nat. Genet.* **44**, 260–268 (2012).
36. Chen, T., Giri, M., Xia, Z., Subedi, Y. N. & Li, Y. abstracts of papers at the seventieth annual meeting of the society of general physiologists: Genetic and animal models for ion channel function in physiology and disease. *J. Gen. Physiol.* **148**, 183–183. <https://doi.org/10.1085/jgp.148.2.183> (2016).
37. Chen, T., Giri, M., Xia, Z., Subedi, Y. N. & Li, Y. Genetic and epigenetic mechanisms of epilepsy: A review. *Neuropsychiatr. Dis. Treat.* **13**, 1841–1859. <https://doi.org/10.2147/ndt.s142032> (2017).
38. Esser, D. *et al.* Interpreting whole genome and exome sequencing data of individual gastric cancer samples. *BMC Genom.* <https://doi.org/10.1186/s12864-017-3895-z> (2017).
39. Rochus, C. M. *et al.* Revealing the selection history of adaptive loci using genome-wide scans for selection: An example from domestic sheep. *BMC Genom.* <https://doi.org/10.1186/s12864-018-4447-x> (2018).
40. Mu, F. *et al.* Structural characterization and association of ovine dickkopf-1 gene with wool production and quality traits in Chinese merino. *Genes* **8**, 400. <https://doi.org/10.3390/genes8120400> (2017).
41. Brito, L. F. *et al.* Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genom.* <https://doi.org/10.1186/s12864-017-3610-0> (2017).
42. Patel, V. B., Zhong, J.-C., Grant, M. B. & Oudit, G. Y. Role of the ace2/angiotensin 1–7 axis of the renin-angiotensin system in heart failure. *Circ. Res.* **118**, 1313–1326 (2016).
43. de Simoni Gouveia, J. J., da Silva, M. V. G. B., Paiva, S. R. & de Oliveira, S. M. P. Identification of selection signatures in livestock species. *Genet. Mol. Biol.* **37**, 330–342. <https://doi.org/10.1590/s1415-47572014000300004> (2014).
44. Lin, Y. *et al.* Genetic variations in CYP17a1, CACNB2 and PLEKHA7 are associated with blood pressure and/or hypertension in the ethnic minority of china. *Atherosclerosis* **219**, 709–714. <https://doi.org/10.1016/j.atherosclerosis.2011.09.006> (2011).
45. Huang, Z., Qi, G., Miller, J. S. & Zheng, S. G. Cd226: An emerging role in immunologic diseases. *Front. Cell Dev. Biol.* **8**, 564. <https://doi.org/10.3389/fcell.2020.00564> (2020).
46. Hansen, T. H. *et al.* A novel loss-of-function variant in the chloride ion channel gene *clcn2* associates with atrial fibrillation. *Sci. Reports* **10**. <https://doi.org/10.1038/s41598-020-58475-9> (2020).
47. Shi, *et al.* Fatty acid elongase 7 (ELOVL7) plays a role in the synthesis of long-chain unsaturated fatty acids in goat mammary epithelial cells. *Animals* **9**, 389. <https://doi.org/10.3390/ani9060389> (2019).
48. Fröb, F. *et al.* Ep400 deficiency in schwann cells causes persistent expression of early developmental regulators and peripheral neuropathy. *Nat. Commun.* <https://doi.org/10.1038/s41467-019-10287-w> (2019).
49. Thompson, N. A. *et al.* Islet-brain1/JNK-interacting protein-1 is required for early embryogenesis in mice. *J. Biol. Chem.* **276**, 27745–27748. <https://doi.org/10.1074/jbc.c100222200> (2001).
50. Testa, U., Castelli, G. & Pelosi, E. Esophageal cancer: Genomic and molecular characterization, stem cell compartment and clonal evolution. *Medicines* **4**, 67. <https://doi.org/10.3390/medicines4030067> (2017).
51. Colomer, C., Marruecos, L., Vert, A., Bigas, A. & Espinosa, L. NF – κ B members left home: NF – κ B-independent roles in cancer. *Biomedicines* **5**, 26. <https://doi.org/10.3390/biomedicines5020026> (2017).
52. Prevot, V. *et al.* Normal female sexual development requires neuregulin–erbB receptor signaling in hypothalamic astrocytes. *J. Neurosci.* **23**, 230–239. <https://doi.org/10.1523/jneurosci.23-01-00230.2003> (2003).
53. Prevot, V., Lomniczi, A., Corfas, G. & Ojeda, S. R. erbB-1 and erbB-4 receptors act in concert to facilitate female sexual development and mature reproductive function. *Endocrinology* **146**, 1465–1472. <https://doi.org/10.1210/en.2004-1146> (2005).
54. Drouilhet, L. *et al.* The highly prolific phenotype of lacauine sheep is associated with an ectopic expression of the b4galnt2 gene within the ovary. *PLoS Genet.* **9**, e1003809. <https://doi.org/10.1371/journal.pgen.1003809> (2013).

55. Zi, X.-D., Hu, L., Lu, J.-Y., Liu, S. & Zheng, Y.-C. Comparison of the sequences and expression levels of genes related to follicular development and atresia between prolific and nonprolific goat breeds. *Veter. Med. Sci.* **6**, 187–195. <https://doi.org/10.1002/vms3.225> (2019).
56. Chun, C. *et al.* The homologous recombination component EEPD1 is required for genome stability in response to developmental stress of vertebrate embryogenesis. *Cell Cycle* **15**, 957–962. <https://doi.org/10.1080/15384101.2016.1151585> (2016).
57. Jodoin, J. N. *et al.* Nuclear-localized asunder regulates cytoplasmic dynein localization via its role in the integrator complex. *Mol. Biol. Cell* **24**, 2954–2965. <https://doi.org/10.1091/mbc.e13-05-0254> (2013).
58. Mihalas, B. P. *et al.* Dynamin 2-dependent endocytosis is essential for mouse oocyte development and fertility. *FASEB J.* **34**, 5162–5177. <https://doi.org/10.1096/fj.201902184r> (2020).
59. Liu, Q. *et al.* Pre-eclampsia is associated with the failure of melanoma cell adhesion molecule (MCAM/CD146) expression by intermediate trophoblast. *Lab. Invest.* **84**, 221–228. <https://doi.org/10.1038/labinvest.3700033> (2003).
60. Bar-Eli, M. Molecular mechanisms of melanoma metastasis. *J. Cell. Physiol.* **173**, 275–278 (1997).
61. Yoshioka, S. Melanoma cell adhesion molecule (MCAM/CD146) is expressed on human luteinizing granulosa cells: Enhancement of its expression by hCG, interleukin-1 and tumour necrosis factor- α . *Mol. Hum. Reprod.* **9**, 311–319. <https://doi.org/10.1093/molehr/gag042> (2003).
62. Yamashita, T. *et al.* Opn5 is a uv-sensitive bistable pigment that couples with gi subtype of g protein. *Proc. Natl. Acad. Sci.* **107**, 22084–22089 (2010).
63. Pipek, R. P., Kolasa, M., Podkowa, D., Kloc, M. & Kubiak, J. Z. Cell adhesion molecules expression pattern indicates that somatic cells arbitrate gonadal sex of differentiating bipotential fetal mouse gonad. *Mech. Dev.* **147**, 17–27. <https://doi.org/10.1016/j.mod.2017.07.001> (2017).
64. Guerra, D. *et al.* Expression of fibroblast growth factor receptors during development and regression of the bovine corpus luteum. *Reprod. Fertil. Dev.* **20**, 659–664 (2008).
65. Stein, Y. Lipoprotein lipase and atherosclerosis. *Atherosclerosis* **170**, 1–9. [https://doi.org/10.1016/s0021-9150\(03\)00014-5](https://doi.org/10.1016/s0021-9150(03)00014-5) (2003).
66. Stoffel, W. *et al.* $\delta 6$ -desaturase (fads2) deficiency unveils the role of $\omega 3$ - and $\omega 6$ -polyunsaturated fatty acids. *EMBO J.* **27**, 2281–2292 (2008).
67. Zhao, J. *et al.* Sphingosine 1-phosphate protects against radiation-induced ovarian injury in female rats—impact on mitochondrial-related genes. *Reprod. Biol. Endocrinol.* <https://doi.org/10.1186/s12958-020-00659-z> (2020).
68. Luna-Nevarez, G. *et al.* Discovery and validation of candidate SNP markers associated to heat stress response in pregnant ewes managed inside a climate-controlled chamber. *Trop. Anim. Health Prod.* **52**, 3457–3466. <https://doi.org/10.1007/s11250-020-02379-3> (2020).
69. Hodžić, A. *et al.* Genetic variation in circadian rhythm genes CLOCK and ARNTL as risk factor for male infertility. *PLoS ONE* **8**, e59220. <https://doi.org/10.1371/journal.pone.0059220> (2013).
70. Teo, W., McGuigan, M. R. & Newton, M. J. The effects of circadian rhythmicity of salivary cortisol and testosterone on maximal isometric force, maximal dynamic force, and power output. *J. Strength Cond. Res.* **25**, 1538–1545. <https://doi.org/10.1519/jsc.0b013e3181da77b0> (2011).
71. Biesiada, L. *et al.* Identification of placental genes linked to selective intrauterine growth restriction (IUGR) in dichorionic twin pregnancies: Gene expression profiling study. *Hum. Genet.* **138**, 649–659. <https://doi.org/10.1007/s00439-019-02016-6> (2019).
72. Dimitri, P. The role of GLIS3 in thyroid disease as part of a multisystem disorder. *Best Pract. Res. Clin. Endocrinol. Metabol.* **31**, 175–182. <https://doi.org/10.1016/j.beem.2017.04.007> (2017).
73. Nogueira, T. C. *et al.* GLIS3, a susceptibility gene for type 1 and type 2 diabetes, modulates pancreatic beta cell apoptosis via regulation of a splice variant of the BH3-only protein bim. *PLoS Genet.* **9**, e1003532. <https://doi.org/10.1371/journal.pgen.1003532> (2013).
74. Sedlmayr, P., Blaschitz, A. & Stocker, R. The role of placental tryptophan catabolism. *Front. Immunol.* **5**, 230. <https://doi.org/10.3389/fimmu.2014.00230> (2014).
75. Drobná, Z. *et al.* Analysis of maternal polymorphisms in arsenic (+3 oxidation state)-methyltransferase AS3MT and fetal sex in relation to arsenic metabolism and infant birth outcomes: Implications for risk analysis. *Reprod. Toxicol.* **61**, 28–38. <https://doi.org/10.1016/j.reprotox.2016.02.017> (2016).
76. Vemaraju, S., Nayak, G., Miller, W. E., Copenhagen, D. R. & Lang, R. A. Fetal stage melanopsin (opn4) and gnaq ($g\alpha q$) signaling regulates vascular development of the eye. *BioRxiv*, 537225. <https://doi.org/10.1101/537225> (2019).
77. Chen, C.-P. *et al.* Prenatal diagnosis of de novo interstitial deletions involving 5q23.1-q23.3 and 18q12.1-q12.3 by array CGH using uncultured amniocytes in a pregnancy with fetal interrupted aortic arch and atrial septal defect. *Gene* **531**, 496–501. <https://doi.org/10.1016/j.gene.2013.09.010> (2013).
78. Olsen, T. R. Two novel aspartoacylase gene (ASPA) missense mutations specific to norwegian and swedish patients with canavan disease. *J. Med. Genet.* **39**, 55e–555. <https://doi.org/10.1136/jmg.39.9.e55> (2002).
79. van Gent, D. C. & van der Burg, M. Non-homologous end-joining, a sticky affair. *Oncogene* **26**, 7731–7740. <https://doi.org/10.1038/sj.onc.1210871> (2007).
80. Dean, M., Hamon, Y. & Chimini, G. The human atp-binding cassette (abc) transporter superfamily. *J. Lipid Res.* **42**, 1007–1017. <https://doi.org/10.22541/au.164001179.92800452/v1> (2021).
81. Kim, J. I., Li, T., Ho, I.-C., Grusby, M. J. & Glimcher, L. H. Requirement for the c-maf transcription factor in crystallin gene regulation and lens development. *Proc. Natl. Acad. Sci.* **96**, 3781–3785. <https://doi.org/10.1073/pnas.96.7.3781> (1999).
82. Huang, H. *et al.* PTEN affects cell size, cell proliferation and apoptosis during drosophila eye development. *Development* **126**, 5365–5372. <https://doi.org/10.1242/dev.126.23.5365> (1999).
83. Bharti, K., Liu, W., Csermely, T., Bertuzzi, S. & Arnheiter, H. Alternative promoter use in eye development: The complex role and regulation of the transcription factor MITF. *Development* **135**, 1169–1178. <https://doi.org/10.1242/dev.014142> (2008).
84. Donovan, K., Alekseev, O., Qi, X., Cho, W. & Azizkhan-Clifford, J. O-GlcNAc modification of transcription factor sp1 mediates hyperglycemia-induced VEGF-a upregulation in retinal cells. *Invest. Ophthalmol. Vis. Sci.* **55**, 7862–7873. <https://doi.org/10.1167/iovs.14-14048> (2014).
85. Kmoch, S. *et al.* Mutations in PNPLA6 are linked to photoreceptor degeneration and various forms of childhood blindness. *Nat. Commun.* <https://doi.org/10.1038/ncomms6614> (2015).
86. Li, S. *et al.* SNHG3 cooperates with ELAVL2 to modulate cell apoptosis and extracellular matrix accumulation by stabilizing SNAI2 in human trabecular meshwork cells under oxidative stress. *Environ. Toxicol.* **36**, 1070–1079. <https://doi.org/10.21203/rs.3.rs-130712/v1> (2021).
87. Davies, H. *et al.* Mutations of the braf gene in human cancer. *Nature* **417**, 949–954 (2002).
88. de Lange, M. J. *et al.* Distribution of GNAQ and GNA11 mutation signatures in uveal melanoma points to a light dependent mutation mechanism. *PLoS ONE* **10**, e0138002. <https://doi.org/10.1371/journal.pone.0138002> (2015).
89. Aoki, H., Hara, A., Motohashi, T. & Kunisada, T. Protective effect of kit signaling for melanocyte stem cells against radiation-induced genotoxic stress. *J. Investig. Dermatol.* **131**, 1906–1915. <https://doi.org/10.1038/jid.2011.148> (2011).
90. Mengel-From, J., Wong, T. H., Morling, N., Rees, J. L. & Jackson, I. J. Genetic determinants of hair and eye colours in the scottish and danish populations. *BMC Genet.* <https://doi.org/10.1186/1471-2156-10-88> (2009).
91. Liang, R., Limesand, S. W. & Anthony, R. V. Structure and transcriptional regulation of the ovine placental lactogen gene. *Eur. J. Biochem.* **265**, 883–895. <https://doi.org/10.1046/j.1432-1327.1999.00790.x> (2001).

92. Deol, K., Varghese, R., Wagner, G. & DiMattia, G. Dynamic regulation of mouse ovarian stanniocalcin expression during gestation and lactation 1. *Endocrinology* **141**, 3412–21. <https://doi.org/10.1210/endo.141.9.7658> (2000).
93. Li, K. *et al.* Identification of STC1 as an β -amyloid activated gene in human brain microvascular endothelial cells using cDNA microarray. *Biochem. Biophys. Res. Commun.* **376**, 399–403. <https://doi.org/10.1016/j.bbrc.2008.08.158> (2008).
94. Block, G. J. *et al.* Multipotent stromal cells are activated to reduce apoptosis in part by upregulation and secretion of stanniocalcin-1. *Stem Cells* **27**, 670–681. <https://doi.org/10.1634/stemcells.2008-0742> (2009).
95. Huang, L. *et al.* Anti-inflammatory and renal protective actions of stanniocalcin-1 in a model of anti-glomerular basement membrane glomerulonephritis. *Am. J. Pathol.* **174**, 1368–1378. <https://doi.org/10.2353/ajpath.2009.080476> (2009).
96. Song, G. *et al.* Stanniocalcin 1 is a luminal epithelial marker for implantation in pigs regulated by progesterone and estradiol. *Endocrinology* **150**, 936–945. <https://doi.org/10.1210/en.2008-1026> (2009).
97. Chang, A.C.-M. *et al.* A novel human cDNA highly homologous to the fish hormone stanniocalcin. *Mol. Cell. Endocrinol.* **112**, 241–247. [https://doi.org/10.1016/0303-7207\(95\)03601-3](https://doi.org/10.1016/0303-7207(95)03601-3) (1995).
98. Varghese, R., Wong, C. K. C., Deol, H., Wagner, G. F. & DiMattia, G. E. Comparative analysis of mammalian stanniocalcin genes this work was supported by grants from london health sciences research, inc. (to g.e.d.), the london regional cancer center (to g.e.d.), and the medical research council of canada (to g.f.w.). *Endocrinology* **139**, 4714–4725. <https://doi.org/10.1210/endo.139.11.6313> (1998).
99. Jepsen, M. R. *et al.* The proteolytic activity of pregnancy-associated plasma protein-a is potentially regulated by stanniocalcin-1 and -2 during human ovarian follicle development. *Hum. Reprod.* **31**, 866–874. <https://doi.org/10.1093/humrep/dew013> (2016).
100. Luo, C.-W., Kawamura, K., Klein, C. & Hsueh, A. J. W. Paracrine regulation of ovarian granulosa cell differentiation by stanniocalcin (STC) 1: Mediation through specific STC1 receptors. *Mol. Endocrinol.* **18**, 2085–2096. <https://doi.org/10.1210/me.2004-0066> (2004).
101. Oonk, R. B., Parker, K. L., Gibson, J. L. & Richards, J. S. Rat cholesterol side-chain cleavage cytochrome p-450 (p-450sc) gene: Structure and regulation by cAMP in vitro. *J. Biol. Chem.* **265**, 22392–22401. [https://doi.org/10.1016/s0021-9258\(18\)45718-3](https://doi.org/10.1016/s0021-9258(18)45718-3) (1990).
102. GOLDRING, N. B. *et al.* Cholesterol side-chain cleavage p450 messenger ribonucleic acid: Evidence for hormonal regulation in rat ovarian follicles and constitutive expression in corpora lutea. *Endocrinology* **120**, 1942–1950. <https://doi.org/10.1210/endo-120-5-1942> (1987).
103. Zlotkin, T., Farkash, Y. & Orly, J. Cell-specific expression of immunoreactive cholesterol side-chain cleavage cytochrome p-450 during follicular development in the rat ovary. *Endocrinology* **119**, 2809–2820. <https://doi.org/10.1210/endo-119-6-2809> (1986).
104. Tamura, K., Yoshie, M., Hashimoto, K. & Tachikawa, E. Inhibitory effect of insulin-like growth factor-binding protein-7 (IGFBP7) on in vitro angiogenesis of vascular endothelial cells in the rat corpus luteum. *J. Reprod. Dev.* **60**, 447–453. <https://doi.org/10.1262/jrd.2014-069> (2014).
105. Li, J. *et al.* Novel insights into the genetic basis of buffalo reproductive performance. *BMC Genom.* **19**, 1–11. <https://doi.org/10.1186/s12864-018-5208-6> (2018).
106. Cheong, A. *et al.* Nuclear-encoded mitochondrial ribosomal proteins are required to initiate gastrulation. *Development* **147**, dev188714. <https://doi.org/10.1242/dev.188714> (2020).

Acknowledgements

The authors gratefully acknowledge the International Sheep Genomics Consortium for permitting the use of the Ovine HapMap genotype data. This study was funded by grants from the National Natural Science Foundation of China (NO: 32060743), Bintuan Science and Technology Program (2022CB001-09) and The Principal Fund of Tarim University (NO: TDZKBS202102).

Author contributions

C.-I.Z. and S.L. designed the experiment. C.-I.Z. conducted the experiments and analyzed the results. C.L. L.Z. Q.C. and S.L. assisted in the analyses of the results and helped in manuscript revision. Z.C. completed the qRT-PCR Assaying sample collection and experiment. C.-I.Z. and S.L. prepared the manuscript. All authors have reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15986-x>.

Correspondence and requests for materials should be addressed to L.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022