Markhor-derived Introgression of a Genomic Region Encompassing *PAPSS2* Confers High-altitude Adaptability in Tibetan Goats

Chao Li,[†,1,2] Yujiang Wu,[†,3] Bingchun Chen,[†,1] Yudong Cai,[†,1] Jiazhong Guo,[†,4] Alexander S. Leonard,[2] Peter Kalds,[1] Shiwei Zhou,[1,5] Jingchen Zhang,[5] Ping Zhou,[6,7] Shangqu Gan,[6,7] Ting Jia,[8] Tianchun Pu,[8] Langda Suo,[3] Yan Li,[1] Ke Zhang,[1] Lan Li,[1] Myagmarsuren Purevdorj,[9] Xihong Wang,[1] Ming Li,[1] Yu Wang,[1] Yao Liu,[1] Shuhong Huang,[1] Tad Sonstegard,[10] Ming-Shan Wang,[11] Stephen Kemp,[12] Hubert Pausch,[2] Yulin Chen,[1] Jian-Lin Han,[*,12,13] Yu Jiang,[*,1] and Xiaolong Wang[*,1]

[1]International Joint Agriculture Research Center for Animal Bio-Breeding, Ministry of Agriculture and Rural Affairs/Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China

[2]Animal Genomics, ETH Zürich, 8092 Zürich, Switzerland

[3]Institute of Animal Science, Tibet Academy of Agricultural and Animal Husbandry Sciences, Lhasa 850009, China

[4]College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China

[5]College of Veterinary Medicine, Northwest A&F University, Yangling 712100, China

[6]Institute of Animal Husbandry and Veterinary Medicine, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi 832000, China

[7]State Key Laboratory of Sheep Genetic Improvement and Healthy Breeding, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi 832000, China

[8]Beijing Key Laboratory of Captive Wildlife Technologies, Beijing Zoo, Beijing 100044, China

[9]Lab of Animal Genetics and Animal Reproductive Technology, Research Institute of Animal Husbandry, Mongolian University of Life Sciences, Ulaanbaatar 17024, Mongolia

[10]Acceligen Inc., St. Paul, MN 55104

[11]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 94720

[12]Livestock Genetics Program, International Livestock Research Institute (ILRI), Nairobi 30709-00100, Kenya

[13]CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: xiaolongwang@nwafu.edu.cn; yu.jiang@nwafu.edu.cn; h.jianlin@cgiar.org.

Associate editor: Dr. Anne Yoder

Article

Abstract

Understanding the genetic mechanism of how animals adapt to extreme conditions is fundamental to determine the relationship between molecular evolution and changing environments. Goat is one of the first domesticated species and has evolved rapidly to adapt to diverse environments, including harsh high-altitude conditions with low temperature and poor oxygen supply but strong ultraviolet radiation. Here, we analyzed 331 genomes of domestic goats and wild caprid species living at varying altitudes (high > 3000 m above sea level and low < 1200 m), along with a reference-guided chromosome-scale assembly (contig-N50: 90.4 Mb) of a female Tibetan goat genome based on PacBio HiFi long reads, to dissect the genetic determinants underlying their adaptation to harsh conditions on the Qinghai-Tibetan Plateau (QTP). Population genomic analyses combined with genome-wide association studies (GWAS) revealed a genomic region harboring the 3′-*phosphoadenosine* 5′-*phosphosulfate synthase* 2 (*PAPSS2*) gene showing strong association with high-altitude adaptability ($P_{GWAS} = 3.62 \times 10^{-25}$) in Tibetan goats. Transcriptomic data from 13 tissues revealed that *PAPSS2* was implicated in hypoxia-related pathways in Tibetan goats. We further verified potential functional role of *PAPSS2* in response to hypoxia in *PAPSS2*-deficient cells. Introgression analyses suggested that the *PAPSS2* haplotype conferring the high-altitude adaptability in Tibetan goats originated from a recent hybridization between goats and a wild caprid species, the markhor (*Capra falconeri*). In conclusion, our results uncover a hitherto unknown contribution of *PAPSS2* to high-altitude adaptability in Tibetan goats on QTP, following interspecific introgression and natural selection.

Key words: **environmental adaptation, high altitude, hypoxia, goat, genomics, genetic introgression.**

Introduction

Goat (*Capra hircus*) is one of the first domesticated livestock species, dating back to ~11,000 years before the present (Zeder 2008; Daly et al. 2021). Goats have expeditiously adapted, at both physiological and genetic levels, to various environments across human civilization and migration (Daly et al. 2018), including challenging agro-ecological conditions like high-altitude environments. Emerging genomic data have been significantly enhancing our understanding of the genetic basis of high-altitude adaptability in humans (Simonson et al. 2010; Huerta-Sánchez et al. 2014; Ouzhuluobu et al. 2020; Quan et al. 2021) as well as domestic and wild animals on the Qinghai-Tibetan Plateau (QTP) (Qiu et al. 2012; Wang et al. 2015, 2016; Hu et al. 2019; Liu et al. 2019; Wu et al. 2020).

Goat pastoralism plays an important role to maintain, utilize, and protect natural resources on QTP. Compared to yak and sheep, goats are able to thrive in regions with harsher environments (Silanikove 2000). Recent studies identified several genes associated with the adaptation of goats to high-altitude environment on QTP, including *DSG3* and *FGF5* (Song et al. 2016; Kumar et al. 2018; Guo et al. 2019, 2020). It is widely accepted that *EPAS1* is a key adaptive gene for Tibetans and domestic animals, including goats, native to high altitudes, suggesting convergent evolution of highlanders on QTP (Wu et al. 2020). However, the adaptability to high altitudes is a complex trait affected by many genetic loci, thereby in-depth examinations of large datasets are needed to dissect novel adaptability-associated genomic regions.

Introgression has emerged as a crucial mechanism to introduce novel genetic variations that may confer adaptively relevant phenotypic consequences. Adaptive introgression has been increasingly reported in humans and many animals native to QTP. Examples include potentially functional introgressions from an extinct hominin group "Denisovans" to Tibetans (Huerta-Sánchez et al. 2014; Hackinger et al. 2016; Hu et al. 2017), Tibetan wolves to Mastiff dogs (Miao et al. 2017; von Holdt et al. 2017; Signore et al. 2019; Wang et al. 2020), yak to Tibetan cattle (Chen et al. 2018; Wu et al. 2018), and Argali to Tibetan sheep (Hu et al. 2019; Cao et al. 2021; Lv et al. 2022). These studies suggest that genetic introgressions play crucial roles in facilitating rapid environmental adaptation. However, it is still unknown how Tibetan goats acquired such adaptability to successfully occupy QTP for thousands of years (Jeong et al. 2016).

In this study, we analyzed large-scale genomic and transcriptomic data from goats native to QTP and low-altitude regions and assembled a genome of a female Tibetan goat de novo from PacBio HiFi long reads. We identified a genomic region encompassing the *PAPSS2* gene to be under strong selection in Tibetan goats and verified it to have driven their adaptation to hypoxic environments. We showed that the adaptability-associated haplotype encompassing *PAPSS2* was introgressed into modern Tibetan goats from a wild caprid species, markhor (*C. falconeri*), native to the Himalayan highlands and other surrounding areas. This study expanded the gene repertoire of hypoxia adaptation in highland-dwelling mammals and provided new insights into their evolutionary origins.

Results

Genome Sequencing and Population Genetic Structure

Whole-genome sequencing (WGS) of 91 goats, including 87 domestic and four wild goats (*C. sibirica*) (fig. 1*A* and supplementary data S1, Supplementary Material online), yielded 27.55 billion 150-bp paired-end reads (3.68 Tb) at an average sequencing depth of 15.85× (11–33×) per individual. These newly generated data were complemented with WGS data of 175 domestic and 65 wild goats (average sequencing depth of 13.97× [3.88–49.35×]) from previous studies (Alberto et al. 2018; Zhang et al. 2018; Guo et al. 2019; Grossen et al. 2020; Zheng et al. 2020) (fig. 1*A*). These 331 goat genomes were classified into seven groups based on their geographical locations or species of origins, including East Asian high-altitude goats (EAS-H; > 3500 m; $n = 76$), East Asian low-altitude goats (EAS-L; < 1200 m; $n = 80$), European goats (EUR; $n = 18$), African goats (AFR; $n = 46$), South Asian goats (SAS; $n = 9$), Southwest Asian goats (SWA; $n = 33$), and wild caprid species ($n = 69$, including Siberian ibex [*C. sibirica*, $n = 8$], Nubian ibex [*C. nubiana*, $n = 2$], Iberian ibex [*C. pyernaica*, $n = 4$], Alpine ibex [*C. ibex*, $n = 29$], markhor [*C. falconeri*, $n = 3$], and bezoar [*C. aegagrus*, $n = 23$]). A total of 72,660,144 SNPs was identified using GATK v3.7.0 (McKenna et al. 2010). After filtering for minor allele frequency (MAF) less than 0.05 and missing rate less than 0.1, a final set of 21,032,588 biallelic SNPs was used in downstream analyses.

Neighbor-joining (NJ) tree reconstruction using whole-genome SNPs confirmed that all domestic goats clustered according to their geographic origins (fig. 1*B*), and all 76 Tibetan goats clustered into a single clade adjacent to lowland goats from East Asia. An ADMIXTURE analysis with $K = 7$ assumed ancestries supported the phylogeographic pattern (fig. 1*C* and supplementary fig. S1, Supplementary Material online). The separation of the East Asian populations was consistent with their origin along an altitudinal gradient (fig. 1*B* and *C*). Consistent with the NJ tree and ADMIXTURE analysis, principal component analysis (PCA) revealed four major clusters corresponding to the Siberian ibex, Nubian ibex, Iberian ibex, and Alpine ibex together with markhor, bezoar, and domestic goats (fig. 1*D*). Further PCA of domestic goats alone separated individuals from Africa, Europe, and Asia along the top principal components (fig. 1*E*). East Asian goat populations were classified into different sub-groups based on their geographical origins and native distributions (supplementary fig. S2, Supplementary Material online). Notably, six Tibetan goat populations

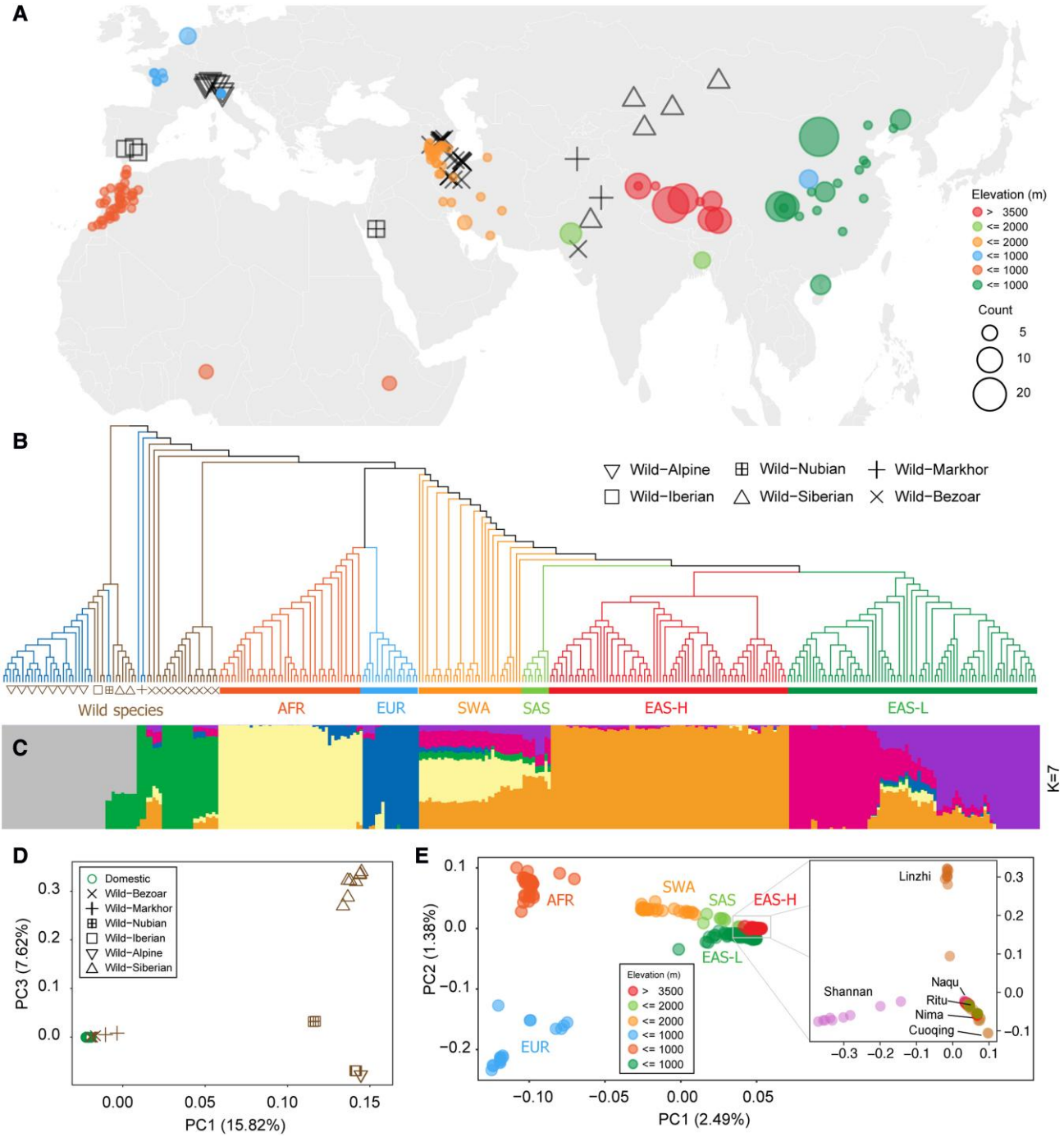

Fig. 1. Geographical distribution and population structure of domestic goats and wild caprid species included in this study. (*A*) Geographical distribution of all domestic goats and wild caprid species. (*B*) NJ phylogenetic tree of 331 goats based on Reynolds' distances estimated using the whole genome-wide SNPs. (*C*) Genetic structure and individual ancestry with colors in each column representing distinct breeds at assumed ancestries of $K = 7$. Plots of principal components 1 and 2 from PCA analysis of all samples (*D*) and domestic goats only (*E*) using the whole genome-wide SNPs. AFR, African goats; EUR, European goats; SWA, Southwest Asian goats; SAS, South Asian goats; EAS-H, East Asian high-altitude goats; and EAS-L, East Asian low-altitude goats.

were divided into three sub-groups following their geographical distributions on QTP (fig. 1*E* and supplementary fig. S2, Supplementary Material online). These findings suggested that Tibetan goats were evolutionarily differentiated from other domestic goats and might have evolved unique genetic features to facilitate their adaptation to high altitudes on QTP.

Genetic Basis of High-altitude Adaptability in Tibetan Goats

To identify genomic regions underlying goat adaptation to high-altitude conditions, we conducted three complementarily comparative genomic analyses between 76 Tibetan and 80 low-altitude goats, that is the fixation index (F_{ST}), the cross-population extended haplotype homozygosity

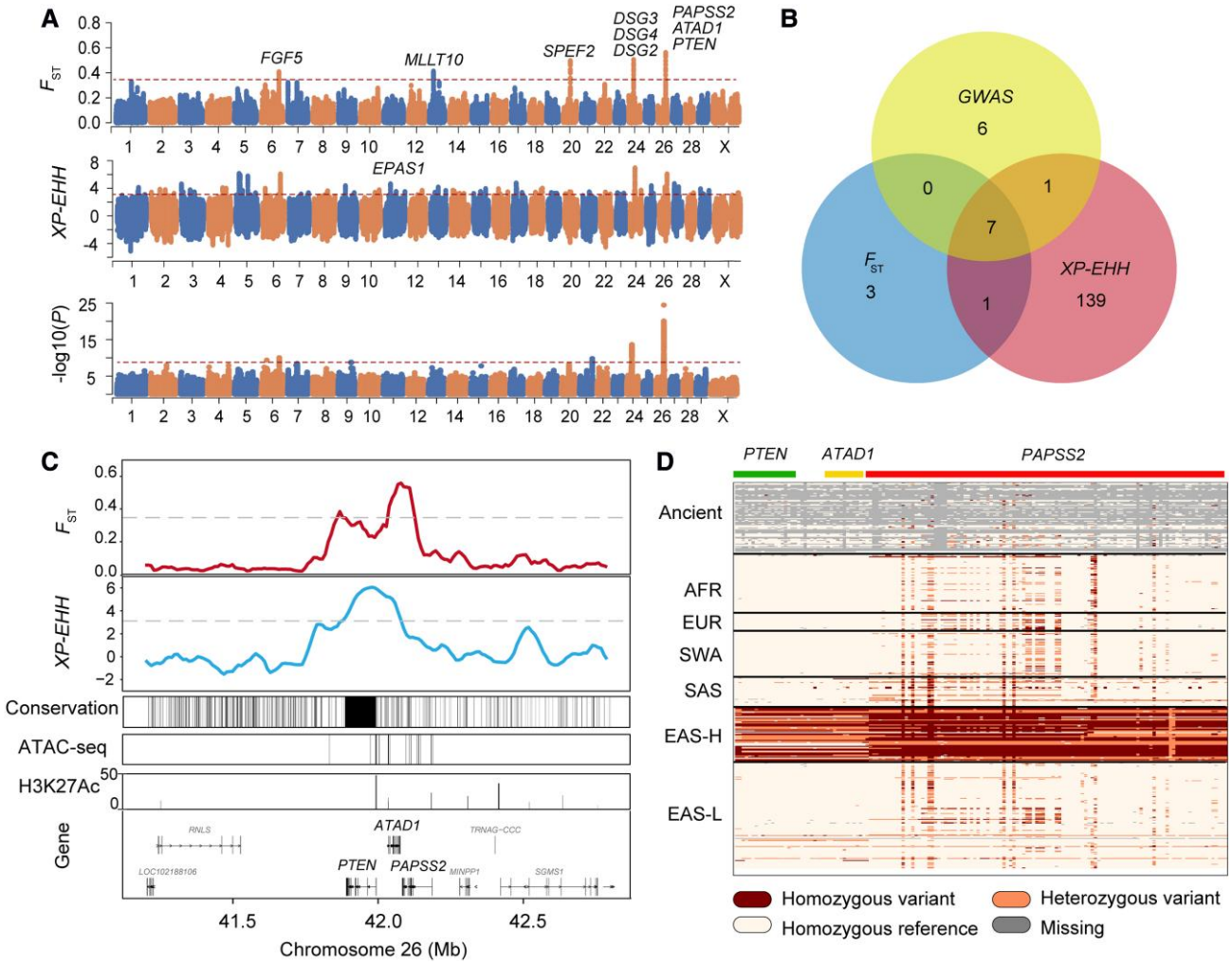

Fig. 2. Genomic regions with strong selection signals and GWAS in Tibetan goats. (*A*) Manhattan plots of genome-wide F_{ST}, *XP-EHH*, and GWAS estimates. The F_{ST} and *XP-EHH* estimates were calculated in 50-kb windows sliding in 10-kb steps along the genomes. The threshold of F_{ST} is 0.36 by permutation. The threshold of *XP-EHH* is 3.15, corresponding to *Z* test and adjusted $P < 0.05$. The threshold of GWAS is $-\log10(P) = 8.65$. (*B*) The number of identified genes with F_{ST}, *XP-EHH*, and GWAS approaches. (*C*) Genomic regions at the *PAPSS2* site in Tibetan and low-altitude goats. (*D*) The pattern of SNP genotypes in the candidate sweep region among all goat samples.

(*XP-EHH*), and a case-control genome-wide association study (GWAS). A sliding window-based approach revealed 50 (F_{ST}) and 551 (*XP-EHH*) putative signatures of selection in genomic regions encompassing 11 (F_{ST}) and 148 (*XP-EHH*) genes (fig. 2*B* and supplementary data S2 and S3, Supplementary Material online). These regions included several genes previously implicated in high-altitude adaptation (*EPAS1* and *DSG3*), hair length (*FGF5*), reproduction (*SPEF2*), and immunity (*MLLT10*) in several species, including goats (Song et al. 2016; Kumar et al. 2018; Cai et al. 2020; Guo et al. 2020; Li et al. 2022; Wu et al. 2020). We further identified a strong signature of selection on chromosome 26 (chr26:41.84–42.14 Mb) encompassing *ATAD1*, *PTEN*, and *PAPSS2*. The GWAS revealed 250 SNPs at five genomic regions that reached genome-wide significance at a Bonferroni-corrected threshold of $P < 2.24 \times 10^{-9}$ (fig. 2*A* and supplementary data S4, Supplementary Material online). These regions encompassed 14 genes, including seven (*ATAD1*, *PTEN*, *PAPSS2*, *FGF5*, *DSG2*, *DSG3*, and *DSG4*) that overlapped those from the F_{ST} and *XP-EHH* analyses (fig. 2*B*), of which the region on chromosome 26 encompassing *ATAD1*, *PTEN*, and *PAPSS2* showed the strongest association ($P = 3.62 \times 10^{-25}$).

The quality, correctness, and completeness of the reference genome as well as the genetic distance between target and reference (e.g., donor of the reference genome) populations can impact population genomic analyses (Lloret-Villas et al. 2021). Analyses of Tibetan goats based on the variants called from the current goat reference genome, known as ARS1, might suffer from reference bias, because the donor of ARS1 was a San Clemente goat (Bickhart et al. 2017), which is likely greatly diverged from Tibetan goats. In order to investigate the utility of a more tailored reference genome to recover high-altitude adaptability-associated genomic regions in goats native to QTP, we assembled a genome of a female Tibetan goat with PacBio HiFi long reads and a reference-guided

approach. This Tibetan goat genome was highly contiguous (contig N50: 90.4 Mb) and exceeded ARS1 in multiple assembly metrics such as correctness and completeness (supplementary data S5, Supplementary Material online). Association and population-genetic analyses (F_{ST}, *XP-EHH*, and GWAS) based on the SNPs called from this Tibetan goat genome mirrored our previous findings that were based on the San Clemente reference sequence (supplementary fig. S3*A*–*C*, Supplementary Material online), with the same strongest signal of selection in the region on chromosome 26 harboring *ATAD1*, *PTEN*, and *PAPSS2*. A new signature of selection in a region on chromosome 3 was nevertheless uncovered in the Tibetan goat genome. The associated region is syntenic with unplaced contigs in ARS1, which was not considered in the initial reference-guided analyses. Association testing with variants called from the ARS1 unplaced contigs recovered a strong association ($P = 2.57 \times 10^{-14}$) of NW_017189568.1:50943 (supplementary fig. S4 and S5, Supplementary Material online). Consistent results from both ARS1 and Tibetan goat genomes as references confirmed that ARS1 is a suitable reference for our downstream analyses. The well-annotated ARS1 reference genome also enabled gene-based analyses that were not readily possible from the draft Tibetan goat assembly.

The most significantly associated SNP (chr26:42,120,872) from the GWAS was in the first intron of *PAPSS2* (fig. 2*C*). This SNP was in strong linkage disequilibrium (LD, $r^2 > 0.8$) with 56 surrounding SNPs (supplementary fig. S6, Supplementary Material online); however, none of these SNPs was in annotated coding or promoter regions. *PAPSS2* encodes PAPS synthase 2, which provides sulfate donor PAPS (3′-phosphoadenosine 5′-phosphosulfate) to all sulfotransferases. Sulfotransferase 1A1 has been determined as a biomarker for high-altitude pulmonary edema resulted from hypoxia-induced changes in the pulmonary circulation (Ahmad et al. 2015). We further examined the genotypes at the most significant 187 SNPs in 542 goats (fig. 2*D*), including the goats sequenced in this study and additional goat genomes (87 ancient goats and 455 domestic goats) retrieved from previous studies (Alberto et al. 2018; Zhang et al. 2018; Guo et al. 2019; Grossen et al. 2020; Zheng et al. 2020). Several alleles in Tibetan goats were either absent or very rare in other goat populations (fig. 2*D*). Within the interval delineated by the 187 SNPs, we detected a 233-bp deletion in intron 7 of *PAPSS2* using Integrative Genomics Viewer (Robinson et al. 2011) (supplementary fig. S7*A* and *B*, Supplementary Material online). The frequency of this homozygous deletion was as high as 80.25% in Tibetan goats, whereas the frequency was 28.75% in the low-altitude goats (supplementary fig. S8, Supplementary Material online). This pattern of allele frequency distribution for the 233-bp deletion in *PAPSS2* was validated by the genotypes of additional 217 goats representing 16 breeds (supplementary fig. S7*C* and *D*, Supplementary Material online), which deserved to examine its potential function.

Global Gene Expression in 13 Tissues of High- and Low-altitude Goats

In addition to these comparative genomic analyses, we conducted transcriptomic analyses of 13 tissues (hypothalamus, cerebellum, cerebrum, heart, liver, spleen, lung, kidney, muscle, rumen, large intestine, small intestine, and skin) from both high- ($n = 4$, Tibetan goats) and low-altitude ($n = 4$, Guanzhong goats) goats. We observed tissue-specific expression in most of the 9,256 differentially expressed genes (DEGs, ranging from 560 in lung to 2,701 in muscle) between the high- and low-altitude goats, which were clustered together in tissues with similar functions (fig. 3*A*, *B* and supplementary data S6, Supplementary Material online). The Gene Ontology (GO) categories of these DEGs were correlated with the functions of specific tissues (fig. 3*C* and supplementary fig. S9, Supplementary Material online). In the six high energy-consuming tissues (heart, liver, spleen, lung, kidney, and muscle), significant GO terms of the DEGs were involved in responses to hypoxia ("blood vessel development" and "muscle structure development"), metabolic process ("generation of precursor metabolites and energy," "organic hydroxy compound metabolic process," "response to lipopolysaccharide," and "regulation of cell activation"), and immune response ("immune effector process," "leukocyte migration," "lymphocyte activation," and "negative/positive regulation of immune system process") (fig. 3*C* and supplementary data S7, Supplementary Material online).

To further investigate the functions of three candidate genes (*PAPSS2*, *PTEN*, and *ATAD1*) within the strongest selective signal on chromosome 26, we compared transcriptomes between the high- and low-altitude goats, and found that *PAPSS2* was differentially expressed in kidney, *PTEN* in both muscle and small intestine, while all three genes were differentially expressed in spleen (supplementary fig. S10*A*, Supplementary Material online). To further validate the expressions of these three genes, we integrated previously published kidney, spleen, and liver transcriptomes (Tang et al. 2017) and found that only *PAPSS2* was differentially expressed in kidney (adjusted $P = 0.0026$) and spleen (adjusted $P = 0.029$) between the high- and low-altitude goats (fig. 3*D*; supplementary fig. S10*B*, Supplementary Material online). Based on the human disease-related gene database (Piñero et al. 2017), further cluster analysis revealed a relationship of the DEGs in goat kidney and spleen with blood vessels and respiratory diseases (supplementary fig. S11*A* and *B*, Supplementary Material online). GO analysis of these DEGs suggested their enrichments in "leukocyte migration and activation", "response to wounding", "PPAR signaling pathway", and immune-related terms (supplementary fig. S11*C*, Supplementary Material online).

Functional Validation of the *PAPSS2* Gene

Given the difference in allele frequency of the 233-bp deletion between the high- and low-land goats, we examined its potential function at the cellular level. We cloned the

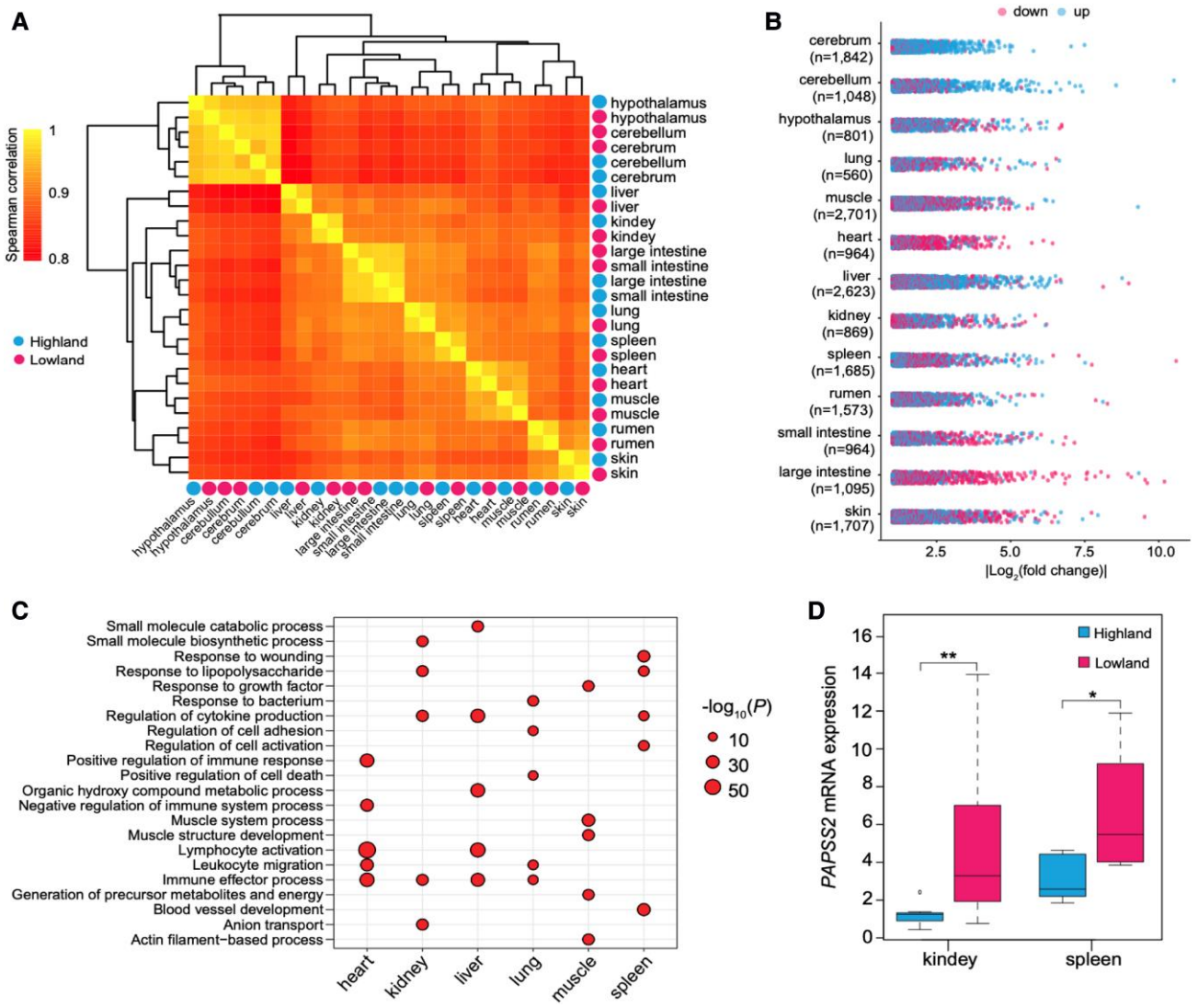

Fig. 3. Global gene expression in 13 tissues between high- and low-altitude breeds. (*A*) Clustering of samples based on whole-genome expression values was calculated as fragments per kilobase million of mRNA (FPKM). The correlation between samples was measured by Spearman's rank correlation coefficient. (*B*) Schematic view of the fold change of differentially expressed genes in 13 tissues (adjusted $P < 0.05$). (*C*) The enriched biological processes in cardiovascular and musculoskeletal tissues. The colored circles in the dot plot represent six tissues and the circle sizes represent the -log10 *P* values with large sizes, indicating a high degree of enrichment. (*D*) *PAPSS2* expression in kidney and spleen. Adjusted $P < 0.05$ (*) and adjusted $P < 0.01$ (**).

deleted fragment from the genomic DNA of a homozygous Tibetan goat into a luciferase reporter vector (pGL3-promoter) and transfected it into HEK293T cells. The result showed no significant change in luciferase expression (Student's *t*-test $P = 0.871$) between the deletion and nondeletion constructs (supplementary fig. S7*E*, Supplementary Material online), suggesting that this deletion did not mediate *PAPSS2* expression in HEK293T cells.

To further explore the functional role of *PAPSS2* in cellular response to hypoxia, we developed a cell model by disrupting *PAPSS2*. We generated a *papss2* knockout cell line ($papss2^{-/-}$) in murine NIH/3T3 cells using the CRISPR/Cas9 system (fig. 4*A* and *B*). Then, we determined the rate of apoptosis in the $papss2^{-/-}$ cells under normoxic (21% O_2) and hypoxic (3% O_2) conditions. In contrast to the apoptosis rate in wild-type cells under normoxia and hypoxia, the $papss2^{-/-}$ cells showed a significantly increased apoptosis rate under hypoxic conditions (fig. 4*C*). A comparative RNA-seq analysis identified 1,660 and 1,380 DEGs between the $papss2^{-/-}$ and wild-type cells under the normoxic and hypoxic conditions, respectively (fig. 4*D*). Interestingly, 349 DEGs were involved in response to hypoxia and oxygen levels, blood vessel morphogenesis and development, circulatory system process, and angiogenesis (fig. 4*E*). These findings suggested a crucial role of *PAPSS2* in blood circulation and thus the adaptive response of Tibetan goats to poor oxygen supply on QTP.

Introgressive Origin of Unique Haplotype of *PAPSS2* in Tibetan Goats From the Markhor

The genomic region encompassing *PAPSS2* showed a significant differentiation between Tibetan and low-altitude

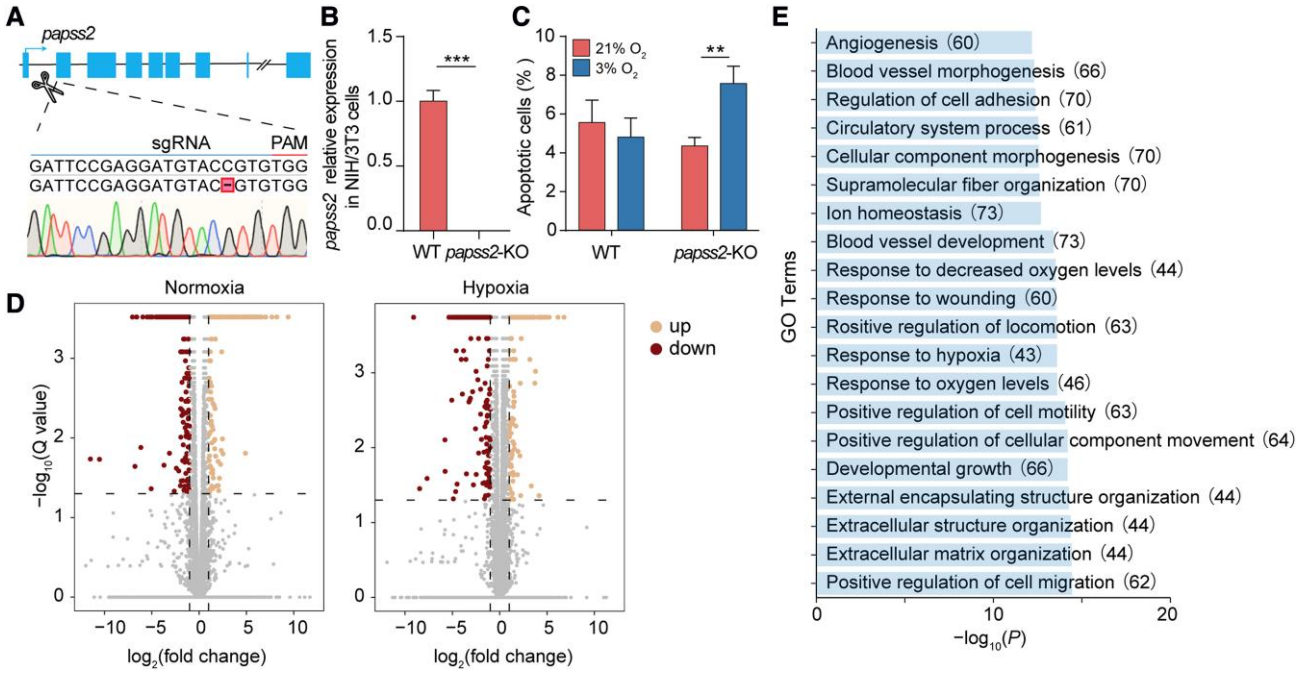

Fig. 4. Functional validation of the *PAPSS2* gene in vitro. (*A*) The edited region of *PAPSS2*. (*B*) The expression of *PAPSS2* in wild-type (WT) and knockout (KO) murine NIH/3T3 cells. ***$P < 0.001$. (*C*) The apoptotic status of the *PAPSS2*-deficient NIH/3T3 cells under hypoxic and normoxic conditions. The cells were assessed after 72 h treatment of hypoxia (3% O_2). **$P < 0.01$. (*D*) Volcano plot of DEGs between the WT and *PAPSS2* KO cells by RNA-seq under normoxic and hypoxic conditions. (*E*) Enriched GO terms using DEGs revealed under the hypoxic condition. The number of enriched genes in each GO term is presented.

goats ($F_{ST} > 0.5$) (fig. 2*C*). Given the fact that Tibetan goats dispersed into QTP in the past few thousand years (Jeong et al., 2016), we hypothesized that the unique alleles of *PAPSS2* in Tibetan goats might have been introgressed from wild caprid species that have evolutionarily adapted to high altitudes. We employed the *D*-statistics analysis to investigate whether the *PAPSS2* alleles of Tibetan goats were derived from a potential interspecific introgression from the markhor and/or four ibex species (fig. 5*A*). This analysis indicated that Tibetan goats shared significantly more alleles with the markhor. We then used the f_d test (Martin et al. 2015) to screen the Tibetan goat genomes for potential introgressions from the markhor based on 50-kb sliding windows. The most significantly introgressed segment (chr26:42,070,001–42,120,000, $f_d = 0.76$; fig. 5*B*) overlapped the genomic region encompassing *PAPSS2*. This potential introgression event was confirmed by Q95 analysis (Racimo et al. 2017) (fig. 5*B*). The mean pairwise sequence divergence (*dxy*) (Martin et al. 2015) in three groups (markhor vs. EAS-H; markhor vs. EAS-L; and EAS-H vs. EAS-L) showed that the *dxy* value in the *PAPSS2* region was significantly reduced in markhor versus EAS-H, but highly elevated in markhor versus EAS-L and EAS-H versus EAS-L (supplementary fig. S12 and S13, Supplementary Material online), which corroborated an introgression event.

Tibetan goats and markhor shared a high similarity in the genomic region encompassing *PAPSS2*, which was clearly different from other domestic goats and wild caprid species (fig. 5*C*). A phylogenetic tree constructed from the 102-kb genomic region encompassing *PAPSS2* confirmed that Tibetan goats and markhor clustered together to form a separate clade, which was divergent from other populations, including those from the East Asian lowland (fig. 5*D* and *E*; supplementary fig. S14 and S15, Supplementary Material online). To further examine the direction of introgression, we performed a partitioned D frequency spectrum (DFS) analysis (Martin and Amos 2021) between EAS-H and the markhor. The positive DFS value was skewed toward low-frequency derived alleles, which suggested a recent gene flow from the markhor to EAS-H (supplementary fig. S16, Supplementary Material online). Collectively, these findings suggested that the interspecific introgression from the markhor, a wild caprid species native to nearby Himalayas, facilitated the rapid adaptation of Tibetan goats to high altitudes on QTP (fig. 5*F*).

Subsequently, we examined if the haplotype (at least 101,703 bp in chr26:42,077,488–42,179,191) shared between Tibetan goats and markhor was a result from incomplete lineage sorting rather than introgression. Using a generation time of 2 years (De Magalhães and Costa, 2009) and a recombination rate of 1.0×10^{-8} (Hu et al., 2019) for goats as well as a divergence time of 1.3 million years ago between goats and markhor (Kumar et al., 2017), the expected length of a shared ancestral tract was only 153.85 bp ($L = 1/(1.0 \times 10^{-8} \times [1.3 \times 10^6/2])$), whereas the probability of the length at 101,703 bp was 1 − GammaCDF (102,000, shape = 2,

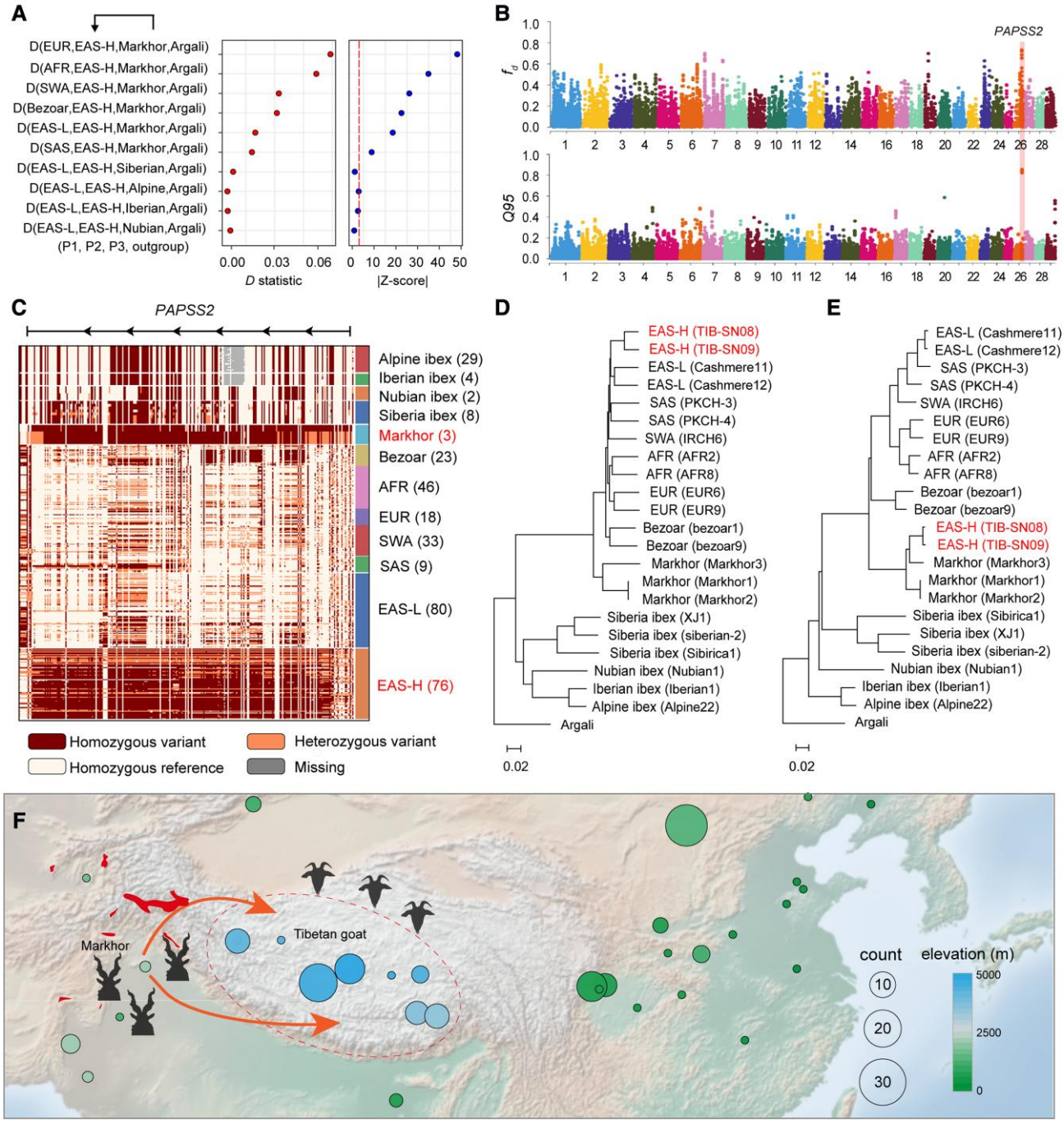

Fig. 5. Identification of genomic region at the *PAPSS2* locus introgressed into Tibetan goats from the markhor. (*A*) Allele sharing between Tibetan goats and wild caprid species. A positive *D* suggests a higher degree of allele sharing between P2 and P3. Statistics with an absolute Z-score greater than 3 are considered to be significant and shown with the red dashed lines. The highest value is observed between EAS-H and the markhor. (*B*) Genome-wide distribution of *D*-statistics (top) and Q95 (bottom) values calculated for 50-kb window sliding with a 10-kb step across the genomes. Each dot represents a 50-kb window. (*C*) The pattern of SNP genotypes in the *PAPSS2* gene region. Rows represent individuals and columns represent SNPs positions in the genomes. NJ phylogenetic tree of the whole-genome sequence (*D*) and the *PAPSS2* region (chr26:42,077,488–42,179,191) (*E*). (*F*) The illustration of genetic introgression from the markhor to Tibetan goats. The geographic distribution of the markhor is according to the International Union for Conservation of Nature (IUCN) Red list marked as the red shaded area (https://www.iucnredlist.org).

rate $= 1/L) = 0$ (supplementary fig. S17, Supplementary Material online). This result suggested that the haplotype shared between Tibetan goats and the markhor was very unlikely due to incomplete lineage sorting but rather a recent interspecific introgression.

Discussion

Unraveling the genetic basis of adaptability in highland mammals improves our understanding of their contemporary evolution of fitness under climate change and provides clues to clinical issues of high-altitude diseases. In this

study, we leveraged 331 whole genomes and 104 transcriptomes from 13 tissues of domestic goats and wild caprid species to determine potential signatures of selection for high-altitude adaptability in Tibetan goats. To address possible reference genome bias, we assembled a highly contiguous and complementary genome of a female Tibetan goat based on PacBio HiFi long reads and a reference-guided approach. We identified *PAPSS2* to show the strongest signal of selection conferring the adaptation of Tibetan goats to high altitudes on QTP, which was likely overlooked in previous studies (Song et al. 2016; Kumar et al. 2018; Guo et al. 2019, 2020). We further revealed that unique alleles of *PAPSS2* in Tibetan goats were derived from a wild caprid species, the markhor, native to Central Asia, Karakoram, and the Himalayas.

The environment on QTP, particularly hypoxic conditions with low oxygen and temperature but strong ultraviolet radiation, poses challenges to both animal and human residents. Understanding the molecular mechanisms involved in the genetic adaptation to harsh conditions could facilitate future breeding of new livestock genotypes that are resilient to changing climate and environmental conditions. In this study, a genomic region on chromosome 26 encompassing *PTEN*, *ATAD1*, and *PAPSS2* was identified as the most striking signal of selection for high-altitude adaptability in Tibetan goats. *ATAD1* plays a critical role in regulating synaptic plasticity, learning, and memory (Ahrens-Nicklas et al. 2017). *PTEN* encodes a lipid dual-specificity phosphatase, which was identified as a tumor suppressor with phosphatase activity. Loss of *PTEN* facilitates HIF-1α-mediated gene expression (Zundel et al. 2000). *PTEN* was also determined as an adaptive gene in Tibetan goats by population genetic analyses (Guo et al. 2019; Jin et al. 2020). However, our complementary analyses involving a more comprehensive set of genomic and transcriptomic data provided strong support for *PAPSS2* to be involved in high-altitude adaptation (fig. 2*A* and supplementary fig. S10, Supplementary Material online). Mutations in *PAPSS2* were reported to be associated with androgen excess (Noordam et al. 2009) and skeletal abnormalities in humans (Miyake et al. 2012; Perez-Garcia et al. 2022). It is suggested that androgen bioactivity is associated with excessive erythrocytosis and chronic mountain sickness (Gonzales and Chaupis 2015). *PAPSS2* was also identified as a major selective gene for high-altitude and cold-environment adaptability in horses (Librado et al. 2015; Shi et al. 2019).

Placing animals with different abilities to adapt to certain environments under uniform conditions may provide clues to understand the target function of selection under environmental change. Such experiments have been conducted, for example, in mice (Cheviron et al., 2014) and chickens (Ho et al., 2020). However, controlled relocation experiments are logistically very difficult to establish for larger animals such as goats. Therefore, we resorted to an in vitro assay to investigate potential contribution of *PAPSS2* to cellular response to hypoxia. Our findings using an in vitro gene knockout assay indicated *PAPSS2* to be functional in the cellular response to hypoxic challenge.

Harsh environments may promote the accumulation of adaptive mutations. It is unlikely that many de novo mutations can emerge at the same genomic region during a short period of evolution (Jeong et al. 2016; Zhao et al. 2017; Hu et al. 2019; Ren et al. 2022). The acquisition of adaptive introgressions from other taxonomically related and spatially close species may contribute to rapid adaptation of recipient species. Recent studies suggest that around 6% of the contemporary Tibetan sheep genomes originate from recent introgression of Argali, which enhanced the capacity of Tibetan sheep to adapt to climatic changes and develop resilience to specific diseases such as pneumonia (Hu et al. 2019; Cao et al. 2021; Lv et al. 2022). Through a series of analyses, we provided additional evidence that Tibetan goats experienced adaptive introgression from the markhor, a wild caprid species native to the Himalayan highlands (fig. 5*C*–*E* and supplementary figs. S13–S16, Supplementary Material online). The most significantly introgressed genomic region overlapped the strongest selection signature in the Tibetan goat genome encompassing *PAPSS2*. The long and unique haplotype of *PAPSS2* (chr26:42,077,488–42,179,191) identified in Tibetan goats also indicated that the haplotype originated from a recent introgression event. Thus, it is plausible that the unique but recent introgression of markhor genes has triggered the adaptability of Tibetan goats to achieve their better fitness to survive over the harsh environment on QTP. In conclusion, we propose that the haplotype introgressed from the markor contributes to an expeditious adaptation of Tibetan goats to high altitudes on QTP. Our findings add to an increasing body of literature that domestic goats, along with other mammals on QTP, underwent strong selection for their rapid adaptation to high altitudes *via* interspecific introgression (Miao et al. 2017; von Holdt et al. 2017; Chen et al. 2018; Wu et al. 2018; Hu et al. 2019; Signore et al. 2019; Wang et al. 2020; Cao et al. 2021; Lv et al. 2022).

Conclusions

Using data from 331 genomes and 104 transcriptomes, we revealed the strongest high-altitude adaptation signature in Tibetan goats at the *PAPSS2* locus. We further demonstrated that the unique haplotype of Tibetan goat *PAPSS2* originated from a recent introgression from the markhor, providing evidence that this interspecific introgression contributed to the high-altitude adaptability in Tibetan goats. Our findings therefore support the important role of interspecific introgression from wild relatives on the rapid adaptation of livestock to extreme environments.

Materials and Methods

Sample Collection and DNA Sequencing

Blood samples from 91 animals (54 from Tibet, 33 from low altitude, and 4 Siberian ibex) were collected for the

isolation of genomic DNA using Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). The DNA was first qualified by 1% agarose gel electrophoresis to monitor its degradation and contamination and then quantified by Qubit® DNA Assay Kit in Qubit® 3.0 Flurometer (Invitrogen, USA). A total amount of 0.2 μg of the DNA per sample passed quality check was used for DNA library preparations. Sequencing library was generated using NEB Next® UltraTM DNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations and index codes were added to each sample. Briefly, the DNA was fragmented by sonication (Covaris, MA, USA) to a size around 350 bp. Then DNA fragments were endpolished, A-tailed, and ligated with the full-length adapter for Illumina sequencing, followed by further PCR amplification. After PCR products were purified by AMPure XP system (Beckman Coulter, Beverly, USA), DNA concentration was measured by the Qubit®3.0 Flurometer, and libraries were analyzed for size distribution by the NGS3K/Caliper and quantified by real-time PCR. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using Illumina PE Cluster Kit (Illumina, USA) according to the manufacturer's instructions. After cluster generation, the DNA libraries were sequenced with Illumina TruSeq SBS Kit v3 on HiSeq 2500 System at Novogene (http://www.novogene.com).

Long Read Sequencing and Assembly of a Tibetan Goat Genome

To provide a high-quality Tibetan goat reference genome, the DNA of a female Tibetan goat sampled in Lhasa of Tibet, China, was used for the PacBio long-read sequencing and genome assembly.

The library was constructed and sequenced on PacBio® RS II System at BGI, Shenzhen, China. HiFi reads were generated based on circular consensus sequencing (CCS). A total of 150 G data corresponding to more than 50-fold genome coverage was generated.

The HiFi reads were assembled with hifiasm v0.16.1 using default parameters, followed by converting the primary contig gfa file into fasta format with gfatools. The contigs were scaffolded using the reference-guided approach of RagTag v2.1.0 with the domestic goat reference genome (the ARS1 assembly, version GCF_001704415.1) (Bickhart et al. 2017), using minimap2 parameters "-cxasm5." We assessed genome quality using Merqury v1.3 (Rhie et al. 2020) with the Illumina short reads from the same individual. We assessed genome completeness using BUSCO v.5.3.2 (Simão et al. 2015) with the "cetartiodactyla" OrthoDB v10 using the metaeuk backend (commit aa7ac2e). We assessed repetitive regions using RepeatMasker (http://www.repeatmasker.org).

Read Alignment and Variant Calling

Trimmomatic v.0.36 (Bolger et al. 2014) was used for quality and adapter trimming before mapping reads to either the domestic goat reference genome (the ARS1 assembly, version GCF_001704415.1) (Bickhart et al. 2017) or the HiFi-based Tibetan goat assembly using BWA-MEM v0.7.15 (Heng 2013). The aligned BAM files were coordinate-sorted, and read groups were merged into the same sample. Duplicates were marked using Picard v2.1 (https://broadinstitute.github.io/picard/). We estimated the coverage distribution at each called site for each sample using QualiMap v2.2 (Okonechnikov et al. 2016). Genotype calling was performed using the HaplotyperCaller and GenotypeGVCF modules of Genome Analysis Toolkit (GATK, v3.7.0) (McKenna et al. 2010). VariantFiltration of GATK was used to remove SNPs with QD < 2.0, FS > 60.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, SOR > 3.0, and MQ < 40.0. Only biallelic variants were retained and then SNPs with MAF less than 0.05 and no more than 10% missing data were extracted using VCFtools v0.1.15 (Danecek et al. 2011) with "–max-missing 0.9 –maf 0.05."

Population Genetic Analysis

For phylogenetic reconstruction, genetic distances were calculated using Plink v1.9 (Purcell et al. 2007). The distance matrix was used to construct an NJ tree implemented in MEGA v6.0 (Tamura et al. 2013). The final tree topology was visualized using iTOL (Letunic and Bork 2016). PCA (Patterson et al. 2006) was performed using smartpca in EIGENSOFT package with default parameters and the following settings: numoutlieriter = 0 and numchrom = 29. Population structure analysis and individual clustering were carried out using ADMIXTURE v1.3 (Alexander et al. 2009), considering between two and eight ancestral clusters.

Identification of Selection Signatures and GWAS

The F_{ST} (Weir and Cockerham 1984) and *XP-EHH* (Sabeti et al. 2007) statistics were calculated between high- and low-altitude goats using VCFtools with a 50-kb sliding window and a step of 10 kb. The threshold of F_{ST} was determined by permutation. The phenotype data were permuted, F_{ST} calculated, and the genome-wide highest F_{ST} value was recorded. We repeated 200 times and ordered the highest 200 values, then we took the lowest value of the top 5% set as the threshold for genome-wide significance. The threshold of *XP-EHH* was determined by the *Z*-transformation of each value and the *Z*-score > 3.7 (corresponding to adjusted $P < 0.05$) was applied to retrieve putative selective sweeps. The intersection of candidate genes obtained by the two methods was used. GWAS was conducted with the genome-wide efficient mixed-model association (GEMMA) software (Zhou and Stephens, 2012).

To investigate the genotypes of candidate regions in other samples, a total of 542 samples were used, including the goats sequenced in this study and additional goat genomes (from other ongoing studies).

RNA-seq

Total RNA of 13 tissues (hypothalamus, cerebrum, cerebellum, heart, liver, spleen, lung, kidney, rumen, large intestine, small intestine, muscle, and skin) from eight adult goats (Tibetan goats at high altitude, $n = 4$; Guanzhong goats at low altitude, $n = 4$) as well as the murine NIH/3T3 cells (wild-type and *papss2*$^{-/-}$ knockout) was extracted using TRIzol reagent from Invitrogen (Thermo Fisher Scientific, MA, USA) according to the manufacturer's instructions and subjected to quality control using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, MA, USA), agarose gel electrophoresis, and the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). The library construction and sequencing were done at the Novogene according to its pipeline (https://www.novogene.com/eu-en/services/research-services/transcriptomesequencing/mrna-sequencing/).

From the raw RNA-seq data, the sequencing adapter reads with unknown nucleotides larger than 5% and the bases with low quality (more than half of the bases' qualities were less than 10) were removed with Trimmomatic v0.36 (Bolger et al. 2014). The clean data were mapped to the goat reference genome (version GCF_001704415.1) using STAR v.020201 (Dobin et al. 2013), and DEGs were calculated using Cufflinks (Trapnell et al. 2012). GO enrichment analysis was conducted using Metascape (http://metascape.org) (Zhou et al. 2019), and R language was used to display significantly enriched GO terms.

Luciferase Reporter Assay

Fragments with and without the *PAPSS2* 233-bp deletion region were amplified through PCR. Primers were listed in supplementary data S8, Supplementary Material online. Then, the fragments were cloned into pGL3-Promoter vector (Miaoling Plasmid Sharing Platform, Wuhan, China) within XmaI site and verified by Sanger sequencing. The HEK293T cells were cultured in DMEM high glucose medium (HyClone) supplied with sodium pyruvate, 10% FBS, 4.0 mM glutamine, streptomycin (100 mg/mL), and penicillin (100 U/mL) at 37 °C in 5% CO_2. The HEK293T cells were transfected with the indicated reporter plasmids together with pRL Renilla Luciferase Control Reporter Vectors (Promega, WI, USA) by Lipofectamine™ 3000 transfection reagent from Invitrogen (Thermo Fisher Scientific, MA, USA). The luciferase activity was measured after 48 h of the transfection with Dual Luciferase Reporter Gene Assay Kit (Beyotime, Shanghai, China).

Genotyping Analysis of the 233-bp Deletion

Genomic DNA was extracted from goat blood samples with Blood Genomic DNA Mini Kit (CWBIO, MA, USA) and used for genotyping and sequencing. To genotype the 233-bp deletion variant, we designed primers that span the deletion region for PCR amplification, and Sanger sequencing was performed (supplementary data S8, Supplementary Material online).

Functional Validation of *PAPSS2* in the NIH/3T3 Cells

The single-guide RNAs (sgRNAs) were designed in the coding sequence (CDS) region of the murine *papss2* gene using an online tool (https://zlab.bio/guide-design-resources). The sgRNAs were ligated into pGL3-U6-sgRNA-PGK-GFP plasmid within BsaI site by Solution I (TakaRa, Tokyo, Japan) and confirmed by Sanger sequencing. Each plasmid was co-transfected into the NIH/3T3 cells (ATCC, CRL-1658) with pST1374-NLS-FLAG-Linker-Cas9 by Lipofectamine™ 3000 transfection reagent. The NIH/3T3 cells were cultured in DMEM high glucose medium with 10% FBS at 37 °C in 5% CO_2. After 72 h of transfection, the fluorescent cells were selected by flow cytometer for monoclonal culture. After enlarged culture, gDNA was extracted for PCR and Sanger sequencing. The *papss2* mRNA expression in the NIH/3T3 cells was determined by quantitative PCR (qPCR). The *papss2*-deficient cells were incubated in a humidified atmosphere under normal (21% O_2) and hypoxic (3% O_2) conditions at 37 °C (Heracell™ 150i CO_2 Incubator, Thermo Fisher Scientific, MA, USA).

Flow cytometry was conducted to assess cell apoptosis using Annexin V-FITC/PI Apoptosis Detection Kit (A211–01, Vazyme, China) according to the manufacturer's instructions. Briefly, the cells were collected after treatment, washed with ice-cold PBS, resuspended in 400 μL of binding buffer, then incubated with 3 μL of Annexin V-FITC and 3 μL of PI at room temperature in the dark for 15 min. The samples were analyzed by a flow cytometer within 1 h (CytoFLEX, Beckman, China). All the cells stained positively for Annexin V-FITC were determined as apoptotic cells.

Genetic Introgression Analysis

D-statistics method (Malinsky et al. 2020) was used to determine whether there was an introgression from wild caprid species into Tibetan goats. We used f_d (Martin et al. 2015) to conduct a genome-wide screening of the introgression regions from wild caprid species with a 50-kb sliding window and a step of 10 kb. The Q95 analysis (Racimo et al. 2017) was used to scan the whole genome with a 50-kb sliding window and a step of 10 kb and re-verify the source of introgression in selected regions. We annotated the adaptive introgression at top 1% of the genome intervals. The genetic relationship of this region was evaluated using the phylogenetic tree by Plink v1.9 (Purcell et al. 2007) with parameters –indep-pairwise 50 5 0.2, while SNPs were filtered with parameter –maf 0.01685.

A partitioned DFS analysis (Martin and Amos 2021) was performed to polarize the introgression between EAS-H and markhor. In this analysis, Argali was used as the outgroup. Markhor was used as *P*3 (source population). EAS-H was set as *P*2, while EAS-L was set as *P*1. The frequency spectrum was calculated by genomics_general (https://github.com/simonhmartin/genomics_general), and the number of samples for *P*1 and *P*2 were both downsampled to 10. Then the DFS was computed from empirical data. Furthermore, we also generated DFS plots under

different demographic models using the online tool (Martin and Amos 2021).

Supplementary material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by grants from the National Key R&D Program (2021YFF1000700), the Tibetan R&D Program (XZ202101ZD0001N and XZ202001YD0017C), the National Natural Science Foundation of China (32161143010, 31972526, and 31772571), and the China Agricultural Research System (CARS-39-03). We thank Meenu Bhati and Danang Crysnanto for their helpful discussions. We appreciate the High-Performance Computing (HPC) of NWAFU for providing computing resources. Xiaolong Wang is a Tang Scholar at NWAFU. The Chinese Government's contribution to the CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources in Beijing (2022-YWF-ZX-02) is appreciated.

Author Contributions

X.-L.W., Y.J., H.P., Y.-L.C., and T.S. conceived the study and designed experiments. X.-L.W., C.L., Y.-D.C., H.P., A.L. performed the majority of analyses with the contribution of X.-H.W., M.L., S.H., and Y.W.; X.-L.W., Y.-J.W., J.Z., S.Z., J.G., S.G., P.Z., L.S., Y.L., K.Z., L.L., T.J., T.P., M.P., and C.L. helped in the preparation of the samples. X.-L.W., B.C., and Yao L. performed the experiments. X.-L.W., C.L., P.K., H.P., J.H., M.W., and S.K. wrote the manuscript. All authors discussed the results and approved the manuscript.

Competing Interest

The authors declare no competing financial interest.

Data Availability

The raw sequencing data have been deposited in the NCBI SRA under Bioproject numbers PRJNA662168, PRJNA767055, PRJNA775840, PRJNA791502 and PRJNA847349.

References

Ahmad Y, Sharma NK, Ahmad MF, Sharma M, Garg I, Srivastava M, Bhargava K. 2015. The proteome of hypobaric induced hypoxic lung: insights from temporal proteomic profiling for biomarker discovery. *Sci Rep*. **5**:10681.

Ahrens-Nicklas RC, Umanah GKE, Sondheimer N, Deardorff MA, Wilkens AB, Conlin LK, Santani AB, Nesbitt A, Juulsola J, Ma E, *et al.* 2017. Precision therapy for a new disorder of AMPA receptor recycling due to mutations in ATAD1. *Neurol Genet*. **3**:e130.

Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P, Benjelloun B, Librado P, Biscarini F, Colli L, *et al.* 2018. Convergent genomic signatures of domestication in sheep and goats. *Nat Commun*. **9**:813.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. **19**: 1655–1664.

Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, *et al.* 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet*. **49**:643–650.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. **30**:2114–2120.

Cai Y, Fu W, Cai D, Heller R, Zheng Z, Wen J, Li H, Xiaolong W, Alshawi A, Sun Z, *et al.* 2020. Ancient genomes reveal the evolutionary history and origin of cashmere-producing goats in China. *Mol Biol Evol*. **37**:2099–2109.

Cao Y-H, Xu S-S, Shen M, Chen Z-H, Gao L, Lv F-H, Xie X-L, Wang X-H, Yang H, Liu C-B, *et al.* 2021. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Mol Biol Evol*. **38**:838–855.

Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, Hu S, Huang S, Zhang H, Zheng Z, *et al.* 2018. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun*. **9**:2337.

Cheviron ZA, Connaty AD, McClelland GB, Storz JF. 2014. Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: transcriptomic plasticity and thermogenic performance. *Evolution* **68**:48–62.

Daly KG, Maisano Delser P, Mullin VE, Scheu A, Mattiangeli V, Teasdale MD, Hare AJ, Burger J, Verdugo MP, Collins MJ, *et al.* 2018. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* **361**:85–88.

Daly KG, Mattiangeli V, Hare AJ, Davoudi H, Fathi H, Doost SB, Amiri S, Khazaeli R, Decruyenaere D, Nokandeh J, *et al.* 2021. Herded and hunted goat genomes from the dawn of domestication in the Zagros Mountains. *Proc Natl Acad Sci U S A*. **118**: e2100901118.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158.

De Magalhaes JP, Costa AJ. 2009. A database of vertebrate longevity records and their relation to other life-history traits. *J Evol Biol*. **22**:1770–1774.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21.

Gonzales GF, Chaupis D. 2015. Higher androgen bioactivity is associated with excessive erythrocytosis and chronic mountain sickness in Andean Highlanders: a review. *Andrologia* **47**:729–743.

Grossen C, Guillaume F, Keller LF, Croll D. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nat Commun*. **11**:1001.

Guo J, Zhong J, Li L, Zhong T, Wang L, Song T, Zhang H. 2019. Comparative genome analyses reveal the unique genetic composition and selection signals underlying the phenotypic characteristics of three Chinese domestic goat breeds. *Genet Sel Evol*. **51**:70.

Guo J, Zhong J, Liu GE, Yang L, Li L, Chen G, Song T, Zhang H. 2020. Identification and population genetic analyses of copy number variations in six domestic goat breeds and Bezoar ibexes using next-generation sequencing. *BMC Genomics*. **21**:840.

Hackinger S, Kraaijenbrink T, Xue Y, Mezzavilla M, Asan, van Driem G, Jobling MA, de Knijff P, Tyler-Smith C, Ayub Q. 2016. Wide distribution and altitude correlation of an archaic high-altitude-adaptive *EPAS1* haplotype in the Himalayas. *Hum Genet*. **135**:393–402.

Heng L. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: https://doi.org/10.48550/arXiv.1303.3997.

Ho WC, Li D, Zhu Q, Zhang J. 2020. Phenotypic plasticity as a long-term memory easing readaptations to ancestral environments. *Sci Adv*. **6**:eaba3388.

Hu H, Petousi N, Glusman G, Yu Y, Bohlender R, Tashi T, Downie JM, Roach JC, Cole AM, Lorenzo FR, *et al.* 2017. Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genet*. **13**:e1006675.

Hu X-J, Yang J, Xie X-L, Lv F-H, Cao Y-H, Li W-R, Liu M-J, Wang Y-T, Li J-Q, Liu Y-G, *et al.* 2019. The genome landscape of Tibetan sheep reveals adaptive introgression from Argali and the history of early human settlements on the Qinghai-Tibetan plateau. *Mol Biol Evol*. **36**:283–303.

Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, *et al.* 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**:194–197.

Jeong C, Ozga AT, Witonsky DB, Malmström H, Edlund H, Hofman CA, Hagan RW, Jakobsson M, Lewis CM, Aldenderfer MS, *et al.* 2016. Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc Natl Acad Sci U S A*. **113**:7485–7490.

Jin M, Lu J, Fei X, Lu Z, Quan K, Liu Y, Chu M, Di R, Wei C, Wang H. 2020. Selection signatures analysis reveals genes associated with high-altitude adaptation in Tibetan goats from Nagqu, Tibet. *Animals (Basel)*. **10**:1599.

Kumar C, Song S, Jiang L, He X, Zhao Q, Pu Y, Malhi KK, Kamboh AA, Ma Y. 2018. Sequence characterization of *DSG3* gene to know its role in high-altitude hypoxia adaptation in the Chinese Cashmere goat. *Front Genet*. **9**:553.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. Timetree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. **34**:1812–1819.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. **44**:W242–W345.

Li Y, Song S, Zhang Z, Liu X, Zhang Y, EG, Ma Y, Jiang L. 2022. A deletion variant within the FGF5 gene in goats is associated with gene expression levels and cashmere growth. *Anim Genet*. **53**:657–664.https://doi.org/10.1111/age.13239

Librado P, Der Sarkissian C, Ermini L, Schubert M, Jónsson H, Albrechtsen A, Fumagalli M, Yang MA, Gamba C, Seguin-Orlando A, *et al.* 2015. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc Natl Acad Sci U S A*. **112**:E6889–E6897.

Liu X, Zhang Y, Li Y, Pan J, Wang D, Chen W, Zheng Z, He X, Zhao Q, Pu Y, *et al.* 2019. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol Biol Evol*. **36**:2591–2603.

Lloret-Villas A, Bhati M, Kadri NK, Fries R, Pausch H. 2021. Investigating the impact of reference assembly choice on genomic analyses in a cattle breed. *BMC Genomics* **22**:363.

Lv F-H, Cao Y-H, Liu G-J, Luo L-Y, Lu R, Liu M-J, Li W-R, Zhou P, Wang X-H, Shen M, *et al.* 2022. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression and agronomically important loci. *Mol Biol Evol*. **39**:msab353.

Malinsky M, Matschiner M, Svardal H. 2020. Dsuite - fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour*. **21**:584–595.

Martin SH, Amos W. 2021. Signatures of introgression across the allele frequency spectrum. *Mol Biol Evol*. **38**(2):716–726.

Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol Biol Evol*. **32**:244–257.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **20**:1297–1303.

Miao B, Wang Z, Li Y. 2017. Genomic analysis reveals hypoxia adaptation in the Tibetan Mastiff by introgression of the gray wolf from the Tibetan Plateau. *Mol Biol Evol*. **34**:734–743.

Miyake N, Elcioglu NH, Iida A, Isguven P, Dai J, Murakami N, Takamura K, Cho TJ, Kim OH, Hasegawa T, *et al.* 2012. PAPSS2 mutations cause autosomal recessive brachyolmia. *J Med Genet*. **49**:533–538.

Noordam C, Dhir V, McNelis JC, Schlereth F, Hanley NA, Krone N, Smeitink JA, Smeets R, Sweep FCGJ, Claahsen-van der Grinten HL, *et al.* 2009. Inactivating PAPSS2 mutations in a patient with premature pubarche. *N Engl J Med*. **360**:2310–2318.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**:292–294.

Ouzhuluobu, He Y, Lou H, Cui C, Deng L, Gao Y, Zheng W, Guo Y, Wang X, Ning Z, *et al.* 2020. *De novo* assembly of a Tibetan genome and identification of novel structural variants associated with high-altitude adaptation. *Natl Sci Rev*. **7**:391–402.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet*. **2**:190.

Perez-Garcia EM, Whalen P, Gurtunca N. 2022. Novel inactivating homozygous *PAPSS2* mutation in two siblings with disproportionate short stature. *AACE Clin Case Rep*. **8**:89–92.

Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. 2017. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. **45**:D833–D839.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, *et al.* 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. **81**:559–575.

Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, Cao C, Hu Q, Kim J, Larkin DM, *et al.* 2012. The yak genome and adaptation to life at high altitude. *Nat Publ Gr*. **44**:946–949.

Quan C, Li Y, Liu X, Wang Y, Ping J, Lu Y, Zhou G. 2021. Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol*. **22**:159.

Racimo F, Marnetto D, Huerta-Sánchez E. 2017. Signatures of archaic adaptive introgression in present-day human populations. *Mol Biol Evol*. **34**:296–317.

Ren L, Yang Y, Qiu M, Brunson K, Chen G, Dong G. 2022. Direct dating of the earliest domesticated cattle and caprines in northwestern China reveals the history of pastoralism in the Gansu-Qinghai region. *J Archaeol Sci*. **144**:105627.

Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. **21**:245.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol*. **29**:24–26.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, *et al.* 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**:913–918.

Shi Y, Fan S, Wu M, Zuo Z, Li X, Jiang L, Shen Q, Xu P, Zeng L, Zhou Y, *et al.* 2019. YTHDF1 links hypoxia adaptation and non-small cell lung cancer progression. *Nat Commun*. **10**:4892.

Signore AV, Yang Y-Z, Yang Q-Y, Qin G, Moriyama H, Ge R-L, Storz JF. 2019. Adaptive changes in hemoglobin function in high-altitude Tibetan Canids were derived via gene conversion and introgression. *Mol Biol Evol*. **36**:2227–2237.

Silanikove N. 2000. The physiological basis of adaptation in goats to harsh environments. *Small Rumin Res*. **35**:181–193.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212.

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, *et al.* 2010. Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**:72–75.

Song S, Yao N, Yang M, Liu X, Dong K, Zhao Q, Pu Y, He X, Guan W, Yang N, *et al.* 2016. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). *BMC Genomics* **17**:122.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. **30**:2725–2729.

Tang Q, Gu Y, Zhou X, Jin L, Guan J, Liu R, Li J, Long K, Tian S, Che T, *et al.* 2017. Comparative transcriptomics of 5 high-altitude vertebrates and their low-altitude relatives. *Gigascience*. **6**(12):1–9.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. **7**:562–578.

von Holdt B, Fan Z, Ortega-Del Vecchyo D, Wayne RK. 2017. EPAS1 Variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. *PeerJ*. **5**:e3522.

Wang M-S, Li Y, Peng M-S, Zhong L, Wang Z-J, Li Q-Y, Tu X-L, Dong Y, Zhu C-L, Wang L, *et al.* 2015. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Mol Biol Evol*. **32**:1880–1889.

Wang M-S, Wang S, Li Y, Jhala Y, Thakur M, Otecko NO, Si J-F, Chen H-M, Shapiro B, Nielsen R, *et al.* 2020. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. *Mol Biol Evol*. **37**:2616–2629.

Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, Li Y, Yuan C, Geng R, Lan X, *et al.* 2016. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci Rep*. **6**:38932.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**:1358–1370.

Wu D-D, Ding X-D, Wang S, Wójcik JM, Zhang Y, Tokarska M, Li Y, Wang M-S, Faruque O, Nielsen R, *et al.* 2018. Pervasive introgression facilitated domestication and adaptation in the *Bos species* complex. *Nat Ecol Evol*. **2**:1139–1145.

Wu D-D, Yang C-P, Wang M-S, Dong K-Z, Yan D-W, Hao Z-Q, Fan S-Q, Chu S-Z, Shen Q-S, Jiang L-P, *et al.* 2020. Convergent genomic signatures of high-altitude adaptation among domestic mammals. *Natl Sci Rev*. **7**:952–963.

Zeder MA. 2008. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. *Proc Natl Acad Sci U S A*. **105**(33):11597–11604.

Zhang B, Chang L, Lan X, Asif N, Guan F, Fu D, Li B, Yan C, Zhang H, Zhang X, *et al.* 2018. Genome-wide definition of selective sweeps reveals molecular evidence of trait-driven domestication among elite goat (*Capra species*) breeds for the production of dairy, cashmere, and meat. *Gigascience* **7**:giy105.

Zhao Y-X, Yang J, Lv F-H, Hu X-J, Xie X-L, Zhang M, Li W-R, Liu M-J, Wang Y-T, Li J-Q, *et al.* 2017. Genomic reconstruction of the history of native sheep reveals the peopling patterns of nomads and the expansion of early pastoralism in east Asia. *Mol Biol Evol*. **34**:2380–2395.

Zheng Z, Wang X, Li M, Li Y, Yang Z, Wang X, Pan X, Gong M, Zhang Y, Guo Y, *et al.* 2020. The origin of domestication genes in goats. *Sci Adv*. **6**:eaaz5216.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. **44**:821–824.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. **10**:1523.

Zundel W, Schindler C, Haas-Kogan D, Koong A, Kaper F, Chen E, Gottschalk AR, Ryan HE, Johnson RS, Jefferson AB, *et al.* 2000. Loss of *PTEN* facilitates *HIF-1*-mediated gene expression. *Genes Dev*. **14**:391–396.