

Multi-omic Analyses Shed Light on The Genetic Control of High-altitude Adaptation in Sheep

Chao Li ^{1,2,#}, Bingchun Chen ^{1,#}, Suo Langda ^{3,#}, Peng Pu ^{4,#}, Xiaojia Zhu ⁵, Shiwei Zhou ¹, Peter Kalds ¹, Ke Zhang ¹, Meenu Bhati ², Alexander Leonard ², Shuhong Huang ¹, Ran Li ¹, Awang Cuoji ³, Xiran Wang ⁵, Haolin Zhu ⁵, Yujiang Wu ³, Renqin Cuomu ³, Ba Gui ³, Ming Li ⁶, Yutao Wang ⁷, Yan Li ¹, Wenwen Fang ¹, Ting Jia ⁸, Tianchun Pu ⁸, Xiangyu Pan ⁹, Yudong Cai ¹, Chong He ¹⁰, Liming Wang ^{11,12}, Yu Jiang ¹, Jian-Lin Han ^{13,14}, Yulin Chen ¹, Ping Zhou ^{11,12,*}, Hubert Pausch ^{2,*}, Xiaolong Wang ^{1,*}

¹International 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

²Animal Genomics, ETH Zürich, Zürich 8092, Switzerland

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

⁴School of Biological and Pharmaceutical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

⁵College of Life Sciences, Shaanxi Normal University, Xi'an 710119, China

⁶Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, 78457 Konstanz, Germany

⁷College of Life and Geographic Sciences, Kashi University, Kashi 844000, China

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

⁹Department of Medical Research, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China

¹⁰Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs/Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, College of Information Engineering, Northwest A&F University, Yangling 712100, China

¹¹Institute of Animal Husbandry and Veterinary Medicine, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi 832000, China

¹²State Key Laboratory of Sheep Genetic Improvement and Healthy Production, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi 832000, China

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

¹⁴Livestock Genetics Program, International Livestock Research Institute, Nairobi 00100, Kenya

*Corresponding authors: xiaolongwang@nwafu.edu.cn (Wang X), hubert.pausch@usys.ethz.ch (Pausch H), zhpqxqf@163.com (Zhou P).

#Equal contribution.

Handling Editor: Kai Ye

Abstract

Sheep were domesticated in the Fertile Crescent and then spread globally, where they have been encountering various environmental conditions. The Tibetan sheep has adapted to high altitudes on the Qinghai-Tibet Plateau over the past 3000 years. To explore genomic variants associated with high-altitude adaptation in Tibetan sheep, we analyzed Illumina short-reads of 994 whole genomes representing ~ 60 sheep breeds/populations at varied altitudes, PacBio High fidelity (HiFi) reads of 13 breeds, and 96 transcriptomes from 12 sheep organs. Association testing between the inhabited altitudes and 34,298,967 variants was conducted to investigate the genetic architecture of altitude adaptation. Highly accurate HiFi reads were used to complement the current ovine reference assembly at the most significantly associated β -globin locus and to validate the presence of two haplotypes A and B among 13 sheep breeds. The haplotype A carried two homologous gene clusters: (1) *HBE1*, *HBE2*, *HBB*-like, and *HBBC*, and (2) *HBE1*-like, *HBE2*-like, *HBB*-like, and *HBB*; while the haplotype B lacked the first cluster. The high-altitude sheep showed highly frequent or nearly fixed haplotype A, while the low-altitude sheep dominated by haplotype B. We further demonstrated that sheep with haplotype A had an increased hemoglobin-O₂ affinity compared with those carrying haplotype B. Another highly associated genomic region contained the *EGLN1* gene which showed varied expression between high-altitude and low-altitude sheep. Our results provide evidence that the rapid adaptive evolution of advantageous alleles play an important role in facilitating the environmental adaptation of Tibetan sheep.

Key words: Environmental adaptation; High altitude; Hypoxia; Selection signature; Sheep.

Introduction

An agropastoral lifeway that emerged in the late Holocene enabled the permanent human occupation of the Qinghai-Tibet Plateau (QTP) [1,2]. Evidence from ancient DNA and archaeological remains indicated that Tibetan sheep populated on the QTP from its northeastern part ~ 3100 years ago [3–5]. Transhumant pastoralism is common on the QTP, where livestock graze and

migrate closely with humans, providing a wide variety of animal-based products to indigenous Tibetans [6,7]. Sheep are the most prevalent livestock on the QTP (> 20 million) [8], and Tibetan sheep are well adapted to the mountainous and hypoxic environments on the QTP [9].

Human and animal populations evolved convergently to cope with the harsh conditions in high-altitude environments [10].

Received: 17 September 2023; Revised: 9 January 2024; Accepted: 29 February 2024.

© The Author(s) 2024. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Adaptation to high altitudes is a polygenic trait. Whole-genome scans have identified causative sequence variants that confer adaptability through various mechanisms in several species, including humans [11–14], monkeys [15–16], dogs [17–19], pigs [20], cattle [21], chickens [22,23], antelopes [24], and goats [25–28]. Variants in *EPAS1*, *EGLN1*, and *MKL1* are associated with responses to low oxygen conditions at high altitudes [11,13,29]. In contrast, variants in *GNPAT*, *CDT1*, and *MMP3* enhance tanning ability, thus protecting skin from increased ultraviolet (UV) radiation at high altitudes [15,30,31]. In Tibetan sheep, previous studies have revealed signatures of high-altitude adaptation that encompass functional genes associated with hypoxia and UV signaling pathways [e.g., *HBB* and *MITF*, as well as the hypoxia-inducible factor 1 (HIF-1) pathway] [4,32,33]. However, the causative variants and their genetic mechanisms that confer high-altitude adaptability in sheep are still elusive.

Here, we utilized whole-genome sequence data from 924 domestic and 70 wild sheep inhabiting different ecosystems as well as 96 transcriptomes from 12 organs of sheep raised at different altitudes to explore the signatures of high-altitude adaptation. We investigated top-ranking selected signals related to adaptation to high altitudes, explored transcriptional regulatory changes, inferred sources of adaptation, and identified genes or likely causative variants associated with high-altitude adaptation in Tibetan sheep.

Results

Genome sequencing and population genetic structuring

The whole genomes of 76 domestic sheep and one wild sheep (argali, *Ovis ammon*) were sequenced with short paired-end libraries to an average coverage of 21.37× (ranging from 15.23× to 38.49×) (Table S1). We also downloaded complementary data from public repositories and compiled a collection of 924 domestic and 70 wild sheep inhabiting different ecosystems at varying altitudes (Figure 1A; Table S2) [4,34–38]. The wild relative species included bighorn (*Ovis canadensis*, $n = 6$), thinhorn (*Ovis dalli*, $n = 6$), argali (*O. ammon*, $n = 11$), mouflon (*Ovis gmelini*, $n = 33$), urial (*Ovis vignei*, $n = 6$), and snow sheep (*Ovis nivicola*, $n = 8$). The sequence data were aligned to the current domestic sheep reference genome (ARS-UI_Ramb_v2.0, GCA_016772045.1) [39], and variant calling with the Genome Analysis Toolkit (GATK; v4.1.7.0) [40] yielded 141,008,381 single nucleotide polymorphisms (SNPs).

A principal component analysis (PCA) of the sequence variant genotypes grouped the sheep into three major clusters: (1) the bighorn, thinhorn, and snow sheep; (2) argali; and (3) domestic sheep, Asiatic mouflon, European mouflon, and urial (Figure 1B). Another PCA that included only domestic sheep separated most individuals by their geographic origins (Figure 1C). The observed dispersion pattern was supported by the phylogeographic pattern (Figure S1; Table S3) and agreed well with previous studies [32,35]. Asian sheep formed different sub-clusters based on the geographical locations and/or altitudes where they inhabited (Figure S2A). Additionally, the 145 Tibetan sheep representing 11 populations were divided into 3 sub-clusters following their isolated geographical distributions on the QTP (Figure S2B).

Genome-wide selection for high-altitude adaptation

Animal populations that are adapted to high altitudes have evolved unique physiological responses driven by the genetic signatures of selection in their genomes. To characterize such signatures and linked functional variants in Tibetan sheep genomes, we used the habitat altitudes of 450 sheep as quantitative phenotypes in a mixed model-based association study based on 34,298,967 biallelic variants (Figure 2A). Of these variants, 896 [SNPs and insertions and deletions (INDELs) < 50 bp] reached genome-wide significance at a Bonferroni-corrected threshold of $P < 1.46 \times 10^{-9}$ (Figure 2A; Table S4), while most of them ($n = 813/896$, 90.74%) clustered in two genomic regions, one on chromosome 15 ($n = 788$, including four non-synonymous mutations) and another on chromosome 25 ($n = 25$). This observation was corroborated by the cross-population extended haplotype homozygosity (XP-EHH) analysis (Figure 2B) and the multilocus test of allele frequency differentiation (XP-CLR) (Figure 2C).

The region on chromosome 15 encompassed the β -globin locus ($P = 1.59 \times 10^{-16}$) (Figure 2D). In this region, two families of genes encoding two embryonic hemoglobins (Hbs) [Hb subunit ϵ 1 (*HBE1*) and Hb subunit ϵ 2 (*HBE2*), alongside their homologous *HBE1*-like and *HBE2*-like] and three β -Hbs [Hb subunit β -A (*HBB*) alongside its homologous *HBB*-like, *HBB1*-like, and *HBB2*-like; Hb subunit β -C (*HBBC*); and Hb fetal subunit β -F (*HBBF*)] were annotated in the current sheep reference genome (ARS-UI_Ramb_v2.0), which were highly similar to haplotype A at this locus [41]. This region was previously found to be associated with high-altitude adaptation in sheep, based on an earlier version of the sheep reference genome Oar_v4.0 (GCA_000298735.2) assembled from a Texel sheep [4], in which the β -globin locus belonged to the haplotype B that missed the four genes of *HBE1*, *HBE2*, *HBB*-like, and *HBBC* [41]. The region on chromosome 25 contained the Egl-9 family hypoxia-inducible factor 1 (*EGLN1*) gene ($P = 1.41 \times 10^{-15}$) (Figure 2E). Notably, *EGLN1* is involved in high-altitude adaptation in humans [11,42]. The haplotype diversity in these two regions separated the 924 domestic sheep according to their altitudes (Figure 2D and E, Figure S3).

Global gene expression in 12 organs of high-altitude and low-altitude sheep

We investigated transcriptomic data from 12 organs (cerebrum, hypothalamus, lung, muscle, heart, liver, kidney, spleen, rumen, small intestine, large intestine, and skin) from high-altitude individuals [$n = 4$, Tibetan sheep at 4470 m above sea level (a.s.l.)] and low-altitude individuals ($n = 4$, Hu sheep at 400 m a.s.l.) to identify differentially expressed genes (DEGs). Organ-specific gene regulation was apparent in six organs (heart, liver, spleen, lung, kidney, and muscle) (Figure 3A). There were 4903 DEGs (ranging from 35 in the skin to 3715 in the muscle) between the high-altitude and low-altitude sheep (Figure 3B; Table S5). These DEGs were clustered together in organs with similar functions. Specifically, in the cerebral and hypothalamic tissues, DEGs were predominantly associated with Gene Ontology (GO) terms pertinent to neural functions. Concurrently, in the spleen and skin tissues, the enrichment was predominantly observed within immune-responsive terms. Furthermore, the rumen, large intestine, small intestine, and liver displayed a

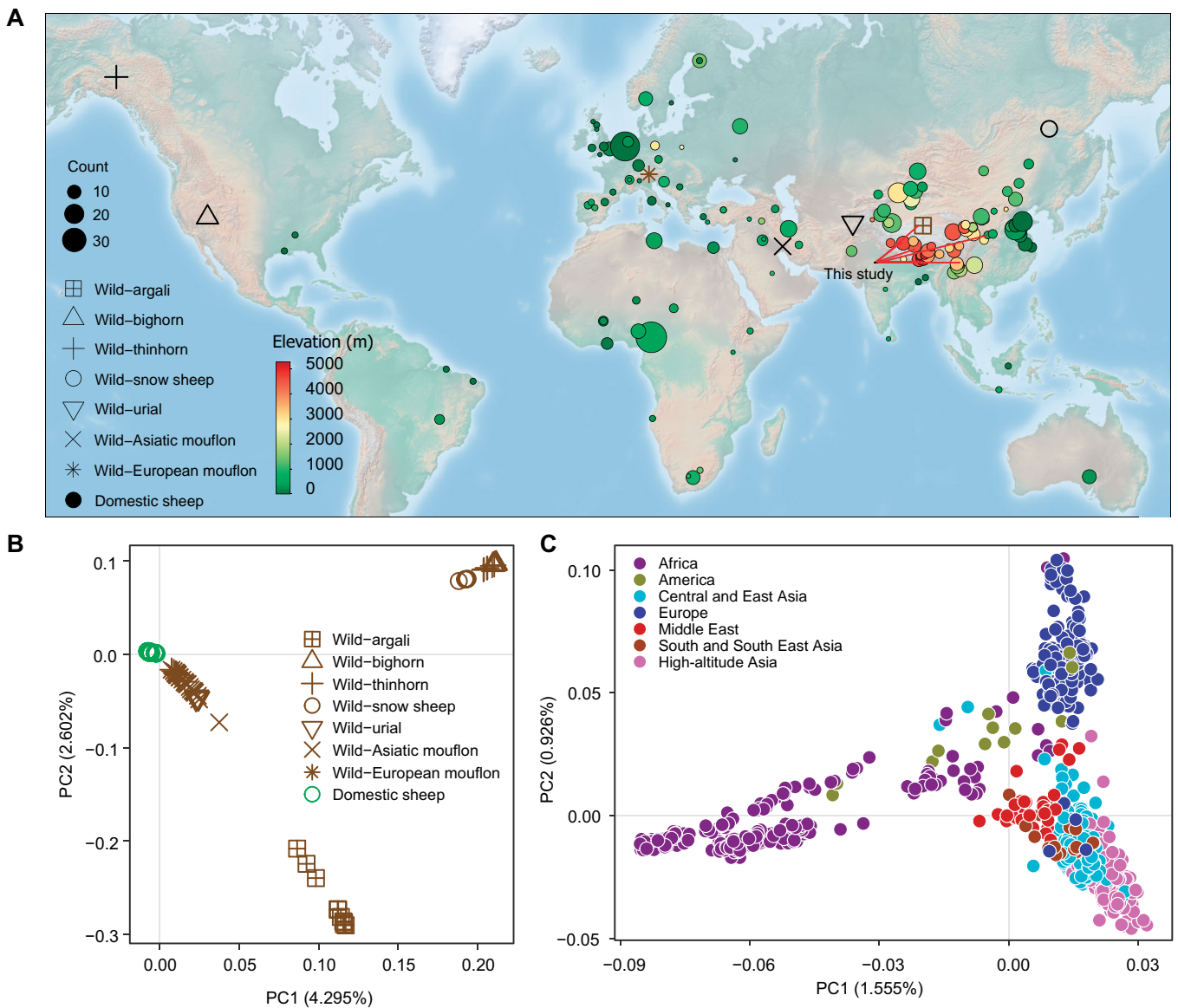


Figure 1 Geographical distribution and genetic diversity of 924 domestic and 70 wild sheep used in this study

A. Geographical distribution of all sheep used in this study. The size of the solid circle represents the number of domestic samples, while the shape indicates the distribution of wild individuals in a geographical region. Red lines indicate the locations of newly sequenced samples. **B.** and **C.** PCA of all samples (**B.**) and domestic sheep only (**C.**). PCA, principal component analysis; PC, principal component.

preponderance of DEGs associated with digestive system-related processes (Table S6).

Significant GO terms of the DEGs in six energy-demanding organs (heart, liver, spleen, lung, kidney, and muscle) were involved in responses to the movement process (“skeletal system morphogenesis”, “response to hormone”, “regulation of tissue remodeling”, and “inorganic ion homeostasis”), metabolic process (“response to nutrient levels”, “regulation of proteolysis”, “organic acid transport”, “carbohydrate metabolic process”, and “aminoglycan catabolic process”), and immune response (“regulation of interleukin-8 production”, “negative regulation of leukocyte activation”, and “inflammatory response”) (Figure 3C; Table S6).

Three (*HBBC*, *HBB*, and *EGLN1*) of the ten genes within the adaptation-associated genomic regions were differentially transcribed (adjusted $P < 0.05$) in at least one organ. The difference in transcript abundance between the high-altitude and low-altitude sheep was pronounced for *HBBC* and *HBB*

(Figure 3D). We further investigated the transcription patterns of these three genes in previously published transcriptomic data of sheep raised at variable altitudes up to 5000 m a.s.l. (<https://www.ncbi.nlm.nih.gov/sra/?term=TIBET+SHEEP>) [43]. These data included 158, 35, and 15 transcriptomes from muscle, lung, and heart, respectively. We found that *HBBC* and *HBB* were highly transcribed in the muscle (*HBB*: $P = 1.09 \times 10^{-6}$; *HBBC*: $P = 7.23 \times 10^{-4}$), lung (*HBB*: $P = 1.62 \times 10^{-5}$; *HBBC*: $P = 4.24 \times 10^{-4}$), and heart (*HBB*: $P = 2.39 \times 10^{-3}$; *HBBC*: $P = 1.34 \times 10^{-3}$) of the high-altitude sheep relative to the low-altitude sheep. On the other hand, the transcription of *EGLN1* was significantly lower in the muscle ($P = 4.36 \times 10^{-4}$) of the high-altitude sheep (Figure 4A). The GO analysis showed that the top 100 DEGs in the heart were enriched in “leukocyte migration”, “regulation of cell activation”, and “wound healing”; the top 100 DEGs in the lung were enriched in “gas transport”, “erythrocytes take up oxygen and release carbon dioxide”, and “inorganic ion transmembrane transport”; and

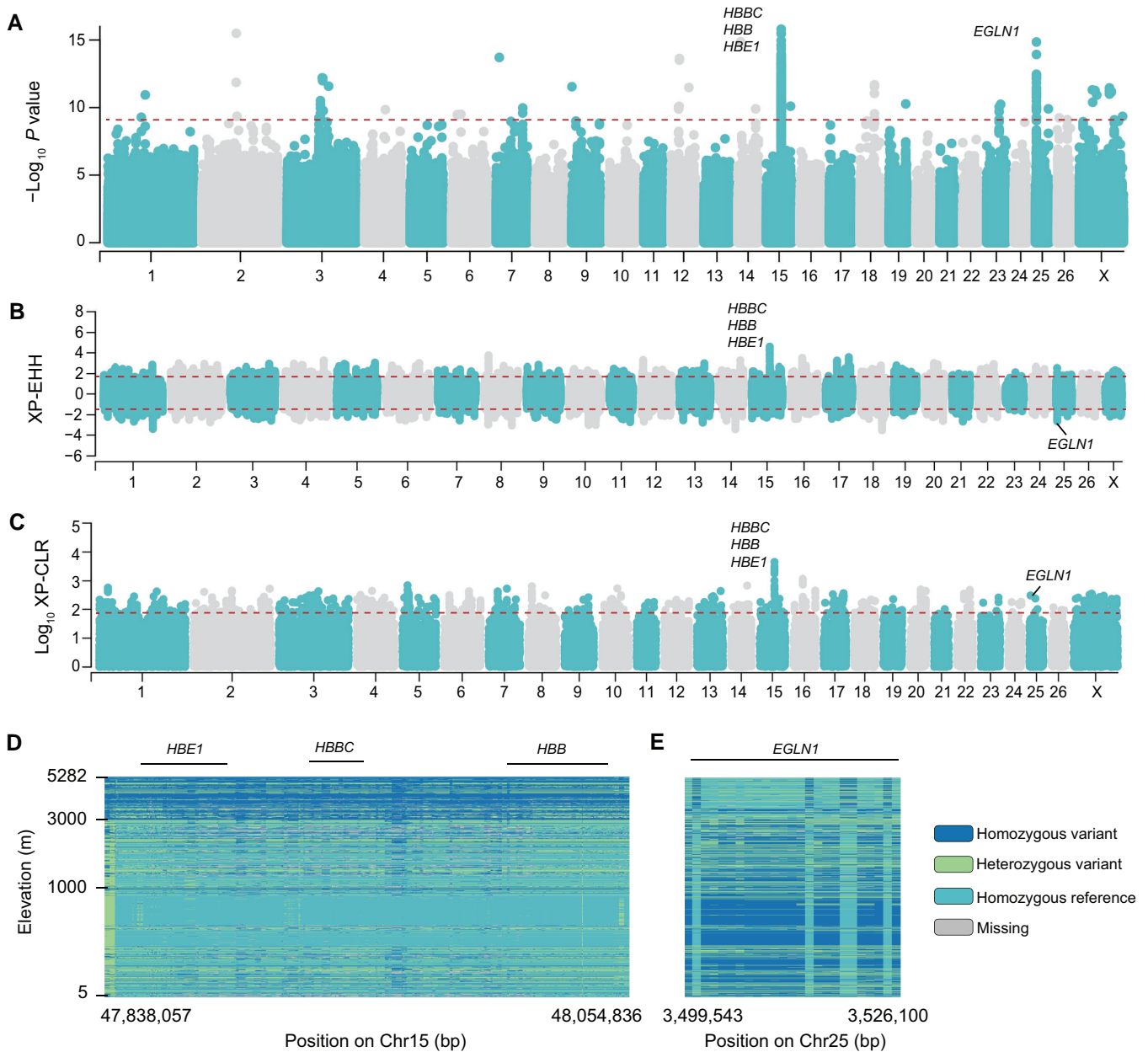


Figure 2 Association analysis between 34,298,967 variants and habitat altitudes in 450 sheep

A. Manhattan plot representing the association of SNPs and INDELs (< 50 bp) with habitat altitudes in sheep ($n = 450$). The horizontal red dotted line indicates the Bonferroni-corrected significance threshold ($-\log_{10} P$ value = 8.84). **B.** and **C.** Manhattan plots of genome-wide XP-EHH (B) and XP-CLR (C) estimates. The XP-EHH and XP-CLR estimates were calculated in 50-kb windows sliding in 10-kb steps along the genomes. The threshold values corresponding to the top 1% of the XP-EHH values ($XP\text{-EHH} < -1.74$ or $XP\text{-EHH} > 1.99$) and the XP-CLR values ($-\log_{10} XP\text{-CLR} > 1.943$) are shown as the horizontal red dotted lines. **D.** and **E.** Comparison of selected genomic regions on Chr15 (D) and Chr25 (E) identified in selection scans of 924 domestic sheep. Chr, chromosome; SNP, single nucleotide polymorphism; INDEL, insertion and deletion; *HBE1*, hemoglobin subunit ϵ 1; *HBB*, hemoglobin subunit β -A; *HBBC*, hemoglobin subunit β -C; *EGLN1*, Egl-9 family hypoxia-inducible factor 1.

the top 100 DEGs in the muscle were enriched in “cellular response to external stimulus”, “response to inorganic substance”, and “ribonucleoprotein complex biogenesis” (Figure 4B).

We collected fresh blood samples from adult sheep living at both high altitudes and low altitudes and measured their blood parameters within 12 h of collection. These records were combined with previously published data [44]. The results showed that Tibetan sheep residing at 4700 m a.s.l. had significantly higher red blood cell (RBC) count ($P = 8.93 \times 10^{-6}$) and Hb concentration ($P = 1.38 \times 10^{-5}$) compared with sheep residing at 500 m a.s.l. (Figure 4C and D). This suggests that Tibetan

sheep augment their oxygen transport capacity by increasing both their RBCs and Hb levels to achieve an enhanced adaptability to high-altitude environments. This observation is consistent with the blood parameters that we previously reported in goats [45].

Mapping potentially functional variants at the β -globin locus

According to the genomic structure at the β -globin locus, domestic sheep were divided into two groups based on the two haplotypes A and B [41]. We examined structural variants

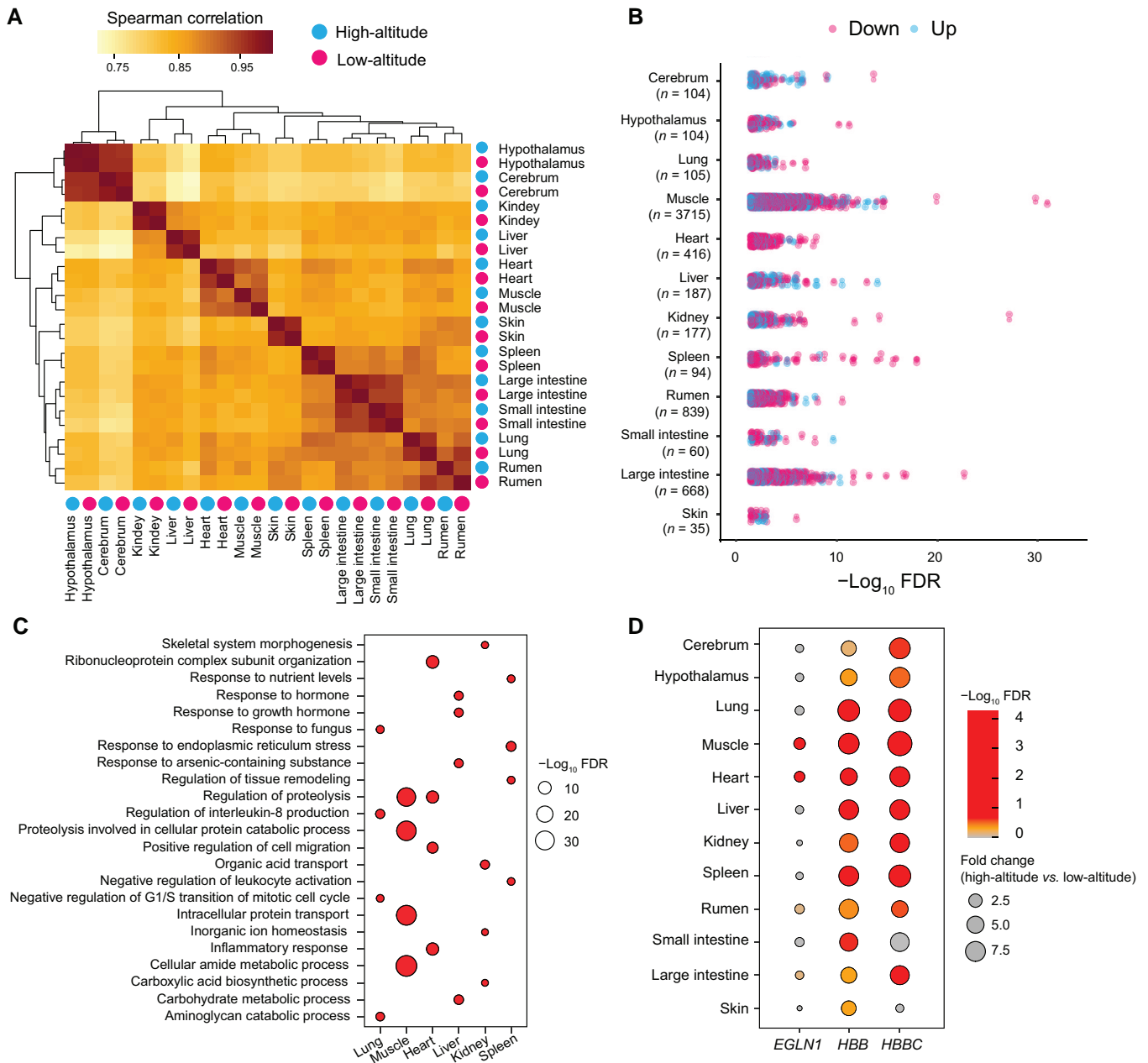


Figure 3 Global gene expression in 12 organs between the high-altitude and low-altitude sheep

A. Clustering of samples was based on whole-genome expression values calculated as the TPM. The correlation between samples was measured by Spearman's rank correlation coefficient. **B.** A schematic view of the DEGs in 12 organs. **C.** The enriched GO biological processes in cardiovascular and musculoskeletal systems. **D.** Global expression of three protein-coding genes with differential expression in at least one organ. TPM, transcripts per million; DEG, differentially expressed gene; GO, Gene Ontology; FDR, false discovery rate.

(SVs) at this locus using the PacBio HiFi data from 13 sheep representing different breeds [46], and found that Tibetan, Yunnan, Merino, East Friesian, Ujumqin, and Charollais sheep carried the haplotype A (Figure 5A), while Kazak, Dorset, Romney, Suffolk, Texel, White Dorper, and Kermani sheep carried the haplotype B (Figure 5A, Figure S4). These findings confirmed that the sheep haplotypes A and B at the β -globin locus were highly diverged, indicating their rather ancient evolutionary origins and likely unique phylogeographic patterns among different sheep breeds/populations [41].

Subsequently, we used short-read sequencing data of 924 domestic sheep to study the distribution pattern of haplotype

A among worldwide populations inhabiting different altitudes. The examination of short-read alignments enabled us to differentiate between the three diplotypes as (1) homozygous haplotype A with high coverage, (2) heterozygous haplotype A with low coverage in the ~ 40 -kb region, and (3) homozygous haplotype B with the large ~ 40 -kb deletion (Figure S5). The distribution of these three diplotypes showed that the frequency of haplotype A increased with altitudes. For instance, its frequency was 0.93 in 98 high-altitude sheep (inhabiting areas around 5000 m a.s.l.), while it was 0.24 in 422 low-altitude sheep (inhabiting areas lower than 1000 m a.s.l.) (Figure 5B). Interestingly, the haplotype A was fixed in

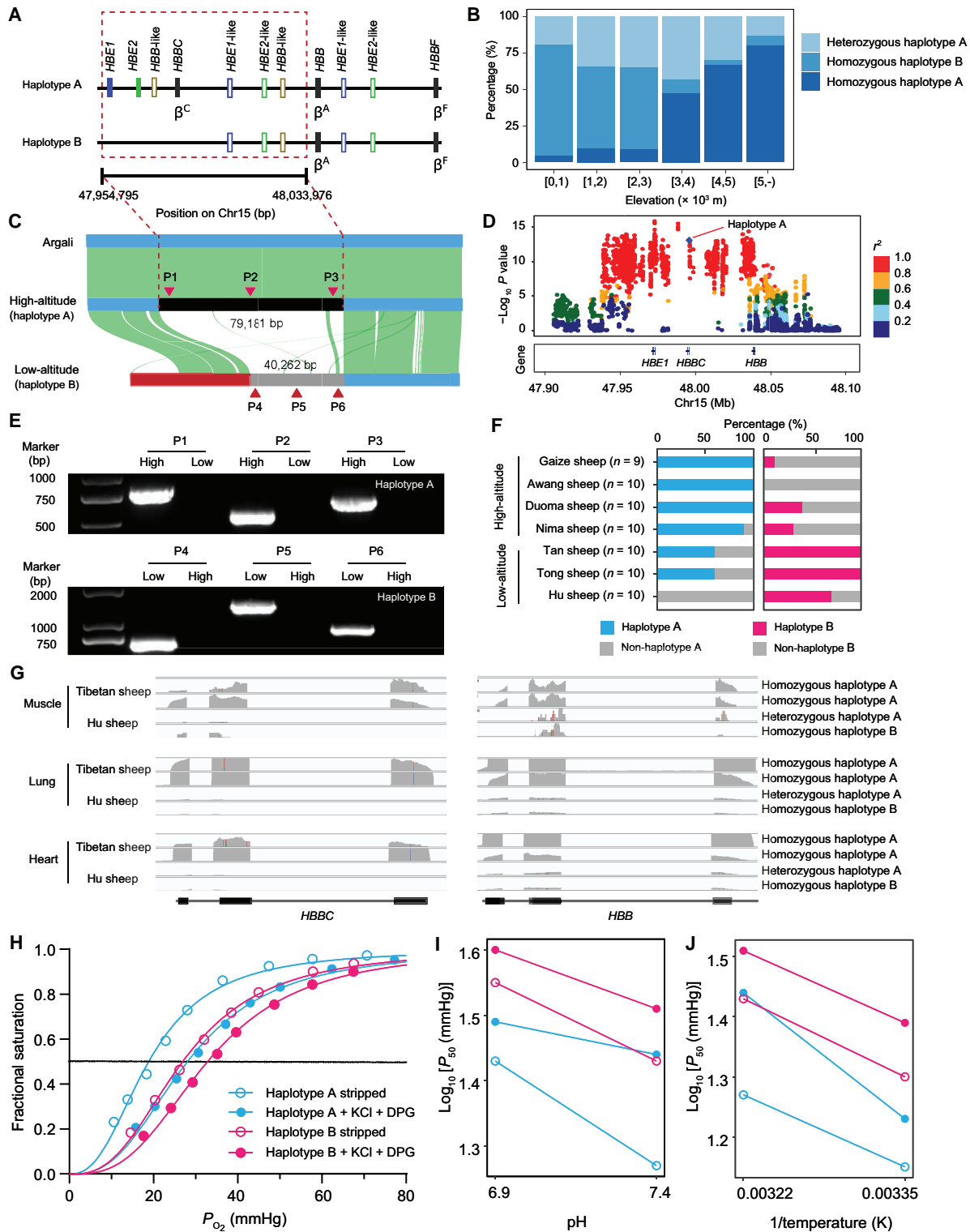


Figure 5 Association of SVs at the β -globin locus with different altitudes

A. A schematic diagram of the sheep β -globin locus. The specific SVs are between haplotypes A and B, where haplotype B possesses an approximately 40-kb deletion. **B.** The frequency distribution of haplotypes A and B in sheep at different altitudes. **C.** Genomic alignment of the β -globin locus in argali, high-altitude (Tibetan), and low-altitude (Texel) sheep. Red triangles indicate primer positions. **D.** The zoom-in GWAS signals on Chr15. LD between the SV (the diamond-shaped blue point) and other variants (SNPs and INDELS < 50 bp) is quantified by the squared Pearson coefficient (r^2). **E.** PCR-based genotyping verified the haplotypes A and B in the high-altitude (High) and low-altitude (Low) sheep. **F.** Frequency of the haplotypes A and B in the high-altitude and low-altitude sheep. **G.** IGW visualization of the mRNA expression between the haplotypes A and B in three organs (muscle, lung, and heart) of Tibetan and Hu sheep. **H.** The O_2 equilibrium curves of haplotypes A and B Hbs in the absence (stripped) or presence of 0.1 M KCl and 0.2 mM DPG at 37°C and pH 7.4. **I.** Bohr effect of haplotypes A and B Hbs at 37°C, as indicated by a plot of $\log_{10} P_{50}$ vs. pH in the absence (stripped) or presence of 0.1 M KCl and 0.2 mM DPG. Bohr factor (Φ) is equal to the slope of the corresponding linear plot. **J.** Temperature effects (reflected by the overall change of enthalpy for oxygenation ΔH°) on the O_2 affinity of haplotypes A and B Hbs at pH 7.4, as indicated by a plot of $\log_{10} P_{50}$ vs. $1/\text{temperature}$ (K) in the absence (stripped) or presence of 0.1 M KCl and 0.2 mM DPG. SV, structural variant; GWAS, genome-wide association study; LD, linkage disequilibrium; PCR, polymerase chain reaction; IGW, Integrative Genomics Viewer; DPG, 2,3-diphosphoglycerate.

their speciation (Figure 5C). The 787 significant SNPs and INDELs within the haplotype A fragment were in strong linkage disequilibrium (LD, $r^2 > 0.8$), suggesting that this fragment has conferred the association signal on chromosome 15 (Figure 5D). Indeed, the signal was absent when the association analysis was conditional on haplotype A (Figure S8). We then evaluated the distribution of haplotypes A and B in additional 69 sheep representing seven breeds (high-altitude: Gaize, Awang, Duoma, and Nima; low-altitude: Tan, Tong, and Hu) using polymerase chain reaction (PCR) amplification (Figure 5E and F; Tables S7–S9). The result confirmed that the haplotype A was dominant in high-altitude sheep, whereas the haplotype B was predominant in low-altitude sheep. Notably, by linking the expression of specific haplotypes of the candidate genes, the transcriptomic data showed that both *HBBC* and *HBB* were highly transcribed in the muscle, lung, and heart of the high-altitude haplotype A carriers, while the transcription of *HBB* was relatively low in the low-altitude haplotype B carriers (Figure 5G).

The O_2 affinity of purified Hbs plays a critical role in realizing the high-altitude adaptability of animals. To determine how blood- O_2 affinity could be affected by the haplotypes A and B, we measured haplotype-specific O_2 -binding properties in purified Hbs. Under different experimental conditions [37°C , $\text{pH} = 7.4$, and the presence or absence of chloride

ions (KCl) and organic phosphates 2,3-diphosphoglycerate (DPG)], the P_{50} of the haplotype A Hbs was significantly smaller than that of the haplotype B ($P < 1.50 \times 10^{-4}$), indicating that the haplotype A had a higher intrinsic O_2 affinity (Figure 5H; Table S10). With the changes in both pH and temperature, the P_{50} values of the haplotypes A and B exhibited broadly similar downward trends (Figure 5I and J). These findings suggest that sheep carrying the high-altitude haplotype A have a significantly enhanced O_2 affinity compared with those carrying the low-altitude haplotype B.

Mapping potentially functional variants in the *EGLN1* region

EGLN1 (also known as HIF prolyl hydroxylase domain 2, *PHD2*) spans ~ 49.4 kb and contains six exons. We detected 23 variants within the *EGLN1* region to be significantly associated with altitudes, including a G/A SNP (Chr25:3,503,284 bp; $P = 1.75 \times 10^{-11}$) in its 3' untranslated region (3' UTR). This SNP was in strong LD ($r^2 > 0.8$) with 22 surrounding SNPs and INDELs (Figure 6A). The allele frequencies of this SNP differed between the Tibetan sheep and low-altitude populations. Up to 77.44% of the Tibetan sheep carried the GG genotype, while only 9.03% of low-altitude sheep had the GG genotype (Figure 6B). The frequency of the G allele was positively correlated with altitudes among the Tibetan sheep and other

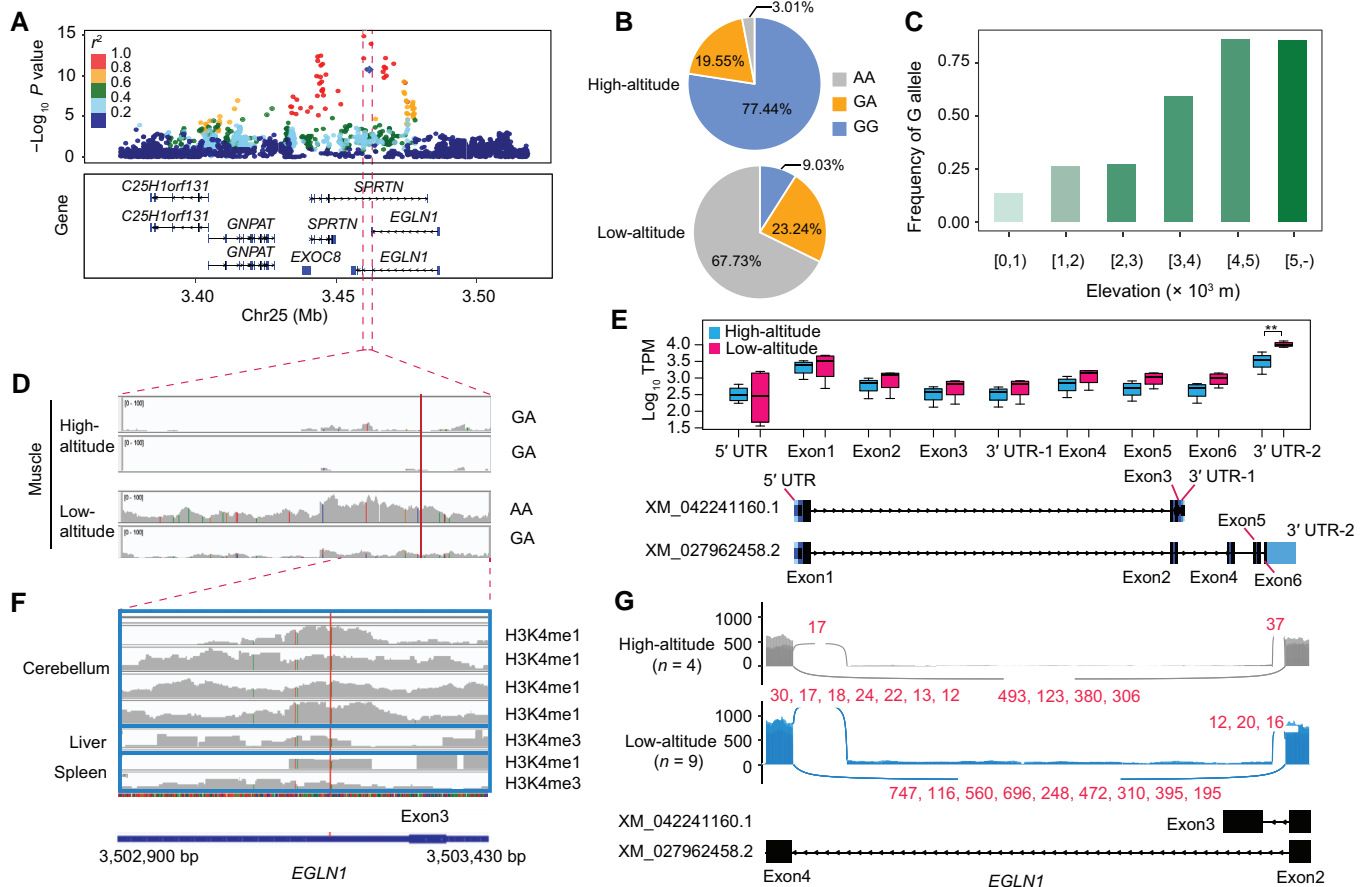


Figure 6 Association of the *EGLN1* G/A SNP (Chr25:3,503,284 bp) with different altitudes

A. The zoom-in GWAS signals on Chr25. LD between the G/A SNP and other variants is quantified by the squared Pearson coefficient (r^2). **B.** Allele frequency of the G/A SNP between the high-altitude and low-altitude sheep. **C.** Allele frequency of the putatively adaptive G allele in worldwide sheep populations. **D.** The expression level of mRNA in the G/A SNP region and its corresponding genotypes. **E.** The exonic expression of *EGLN1* in the muscle between the high-altitude and low-altitude sheep. **F.** The epigenetic signals in the region encompassing the putatively adaptive G allele. **G.** Splice junction patterns between exons 2 and 4. The two transcripts of *EGLN1* at the corresponding positions are shown at the bottom. The numbers in red indicate the reads shared between the two transcripts; each number corresponds to a sample, and the numbers are separated by commas. UTR, untranslated region.

populations ($r^2 = 0.92$) (Figure 6C, Figure S9). The transcriptomic data of the muscle demonstrated that the 3' UTR and all exonic regions of *EGLN1* were down-regulated in high-altitude individuals (Figure 6D and E). We next annotated all 23 significantly associated variants using epigenomic data [47] and found that the G/A SNP was located within a H3K4me1 peak (Figure 6F; Table S4). We further found that an *EGLN1* alternative splicing in the 3' UTR region likely led to an exon elongation in one of the *EGLN1* transcripts, XM_042241160.1 (Figure 6G, Figures S10 and S11). Interestingly, the GC content in the first promoter exon of *EGLN1* was as high as 82.2%. It is known that GC-rich sequences are challenging to sequencing [48,49], but the coverage of mapped DNA reads in this region was surprisingly higher in the high-altitude than in the low-altitude sheep (Figure S12). This suggests that this GC-rich region of the promoter may regulate the activity of *EGLN1*. However, further investigations are required to support such a claim.

Discussion

In this study, we conducted a large-scale genome-wide association scan to elucidate the genetic basis of high-altitude adaptation in sheep. By integrating the large-scale genome-wide association study (GWAS) data with the transcriptomes of 12 organs isolated from high-altitude and low-altitude sheep, we investigated a genomic landscape of high-altitude genetic determinism and highlighted three candidate genes (*HBBC*, *HBB*, and *EGLN1*) and two haplotypes A and B at the β -globin locus associated with high-altitude adaptation in Tibetan sheep.

Along with our recent study that revealed the high-altitude adaptability of Tibetan goats [28], several reports have shown that beneficial alleles facilitated a better fitness of the local dwellers on the QTP. Examples of the adaptive alleles include high-altitude dogs [16,17,50,51], yaks [31], and Tibetan cattle [52,53]. In humans, substantially advantageous alleles enabled Tibetans to be better adapted to environmental challenges such as hypoxia and UV radiation [11–13,30]. The *HBB* locus in Tibetan sheep was previously thought to have been introgressed from argali [4]. However, the genomic region (Chr15:47,400,001–47,500,000 bp) in the Oar_v4.0 (GCA_000298735.2) from a Texel sheep corresponded to the haplotype B [41], in which only *HBE* and *HBB* were annotated [4]. This region was nevertheless recaptured in the Changthangi sheep distributed at high altitude and the variants of its *HBB* and/or neighboring genes were shown as the result of wild introgression, while *HBB*, *HBB*-like, *HBBC*, and *HBE2* were found to have differentiated as putatively selected genes, based on an early sheep reference genome Oar_rambouillet_v1.0 (GCF_002742125.1) [32]. Compared with the complete genomic sequence of the β -globin locus in the latest ARS-UI_Ramb_v2.0 (NC_056068.1:47,967,454–48,084,335) [39], the homologous β -globin locus in the Oar_rambouillet_v1.0 (NC_040266.1:51,913,702–52,096,364) had an extra 66-kb redundant sequence, indicating its poor assembly.

In this study, we utilized the PacBio HiFi reads of 13 domestic sheep [46] and the genome of argali sheep [54] to examine the detailed genomic structure of the β -globin locus, at which the Tibetan sheep and argali carried the haplotype A with highly homologous genomic and gene structures in the ARS-UI_Ramb_v2.0. PCR genotyping of 69 sheep samples validated the dominant distribution of haplotype A in Tibetan sheep, while the genotypes at the β -globin locus of

924 domestic sheep identified a genetic cline of increasing haplotype A alongside the elevated altitudes. Previous studies have indicated that the β -globin locus in sheep contains several SNPs originating from argali through introgression [4]. We also confirmed an extensively phylogeographic distribution of haplotype A in argali, European and Asiatic mouflons, and domestic sheep breeds/populations [41]. All these findings strongly supported ancient and independent evolutionary origins of haplotypes A and B at the β -globin locus in sheep [41], while the haplotype A has been becoming dominant in Tibetan sheep as a result of strong natural selection driving their rapid adaptation to high-altitude challenges. Therefore, we postulated that the haplotype A was likely segregated in the ancestral population of Tibetan sheep before their migration to the QTP. Upon their arrival, the natural selection favoring the haplotype A, in conjunction with the adaptive introgression from argali, may have expedited the fixation of haplotype A within the Tibetan sheep, enabling their rapid adaptation to the high-altitude environment. Moreover, the differential expression of both *HBBC* and *HBB* in the muscle, lung, and heart between high-altitude and low-altitude sheep provided solid evidence that the functional β -globin genes embedded in the haplotype A increased the oxygen transportation of Tibetan sheep to cope with the hypoxic conditions on the QTP [55–57]. Additionally, sheep carrying haplotype A exhibited an enhanced Hb–O₂ affinity, suggesting a more effective bindability to oxygen under hypoxic conditions. In contrast, those with the low-altitude haplotype B had a relatively low Hb–O₂ affinity, allowing their survival in the areas with ample oxygen supply.

We also detected an adaptive allele at the second strongest locus around the *EGLN1* gene, which showed differential expression in muscle between the high-altitude and low-altitude sheep, suggesting its contribution to the adaptation of Tibetan sheep to high-altitude environments. *EGLN1* was also found to be associated with high-altitude adaptation in humans [11,42], cattle [53], and ducks [58], indicating its convergent evolution across different species. *EGLN1* has been reported to be a regulator of *EPAS1*, which directly regulates the expression of erythropoietin (EPO) and eventually influences the Hb level in the blood [42]. Under the hypoxic challenge, hydroxylation significantly decreases, and *EPAS1* and HIF1 α are stabilized, leading to an improved hypoxia adaptation [59]. Although a previous study indicated that *EPAS1* was under selection in Tibetan sheep [33], the results of all three methods (GWAS, XP-EHH, and XP-CLR) applied in this study did not recapture any signals of positive selection for *EPAS1*. Given these observations, we assumed that varying population differences may account for the discrepancies across genome scans. Similar to particular morphological traits such as horn size and shape [4,37], the beneficial allele of *EGLN1* was likely involved in the adaptive evolution of Tibetan sheep to withstand harsh environments on the QTP.

Conclusion

In summary, based on the integrating analysis of short-read sequencing, PacBio HiFi, and transcriptomic data together with the Hb–O₂ affinity of sheep at different altitudes, our results validated that haplotype A was associated with hypoxic adaptation, and further confirmed the ancient and independent evolutionary origins of haplotypes A and B at the β -globin locus. We

also detected a new gene, *EGLN1*, associated with high-altitude adaptability in Tibetan sheep. These data and findings provide valuable resources for livestock breeding and contribute to the understanding of high-altitude diseases in humans.

Materials and methods

Sample collection and DNA sequencing

Blood samples from 77 sheep were collected from different geographic locations: 64 from high-altitude Nima, Kangma, Gaize, Gangba, Duoma, and Awang of Xizang Autonomous Region (> 3000 m a.s.l.), 10 from Lanping, Yunnan Province (~ 1600 m a.s.l.), two from low-altitude Yangling, Shaanxi Province (< 1000 m a.s.l.), and one argali sampled at the Beijing Zoo, Beijing, China. DNA was isolated from whole blood samples using the DNeasy Blood and Tissue Kit (Catalog No. 69506, Qiagen, Dusseldorf, Germany). Paired-end sequencing libraries were prepared using the Genomic DNA Sample Prep Kit (Catalog No. 20060060, Illumina, San Diego, CA). Libraries were sequenced using the Illumina Sequencing Kit v3 (Catalog No. MS-102-3001, Illumina) on an HiSeq 2500 (Catalog No. Hiseq 2500, Illumina) instrument at Novogene (<http://cn.novogene.com>).

Read alignment and variant calling

Trimmomatic (v0.36) [60] was used for quality control and adapter trimming before reads were mapped to the domestic sheep reference genome (ARS-UI_Ramb_v2.0, GCA_016772045.1) [39] using BWA-MEM (v0.7.15) [61]. The aligned Binary Alignment/Map format (BAM) files were coordinate-sorted, and read groups were merged into the same sample using SortSam and MergeSamFiles from the Picard tools software suite (v2.1; <https://broadinstitute.github.io/picard/>). Duplicates were also marked using the Picard tools software suite (v2.1; <https://broadinstitute.github.io/picard/>). The coverage at each called site was estimated for each sample using Qualimap (v2.2) [62]. Genotype calling was performed using HaplotypeCaller and GenotypeGVCFs from the GATK (v4.1.7.0) [40]. The VariantFiltration module of GATK was used to remove SNPs if $QD < 2.0$, $FS > 60.0$, $MQRankSum < -12.5$, $ReadPosRankSum < -8.0$, $SOR > 3.0$, and $MQ < 40.0$. INDELs up to 50 bp were also retained and filtered using $maxIndelSize = 50$, $QD < 2.0$, $FS > 200.0$, and $ReadPosRankSum < -20.0$.

Population genetic analysis

PLINK (v1.9) [63] was used to retain biallelic SNPs that had a minor allele frequency (MAF) higher than 0.05 and did not show an excess of heterozygosity (--hwe 0.001). We excluded SNPs in LD with $r^2 > 0.2$ (--indep-pairwise 50 5 0.2). These filters retained 4,899,522 high-quality SNPs. A PCA was performed with the smartpca program from the EIGENSOFT package [64] with default parameters and the following settings: numoutlieriter = 0 and numchrom = 26.

Association tests

To exclude the effect of tail size, we only used thin-tailed sheep similar to Tibetan sheep for altitude association analysis and obtained a set of 450 individuals from various populations and altitudes. BCFtools [65] was used to extract 450 individuals from the total unfiltered Variant Call Format (VCF) file, and PLINK (v1.9) [63] was used to extract 34,298,967 variants (31,567,131 SNPs and 2,731,836

INDELs) with MAF higher than 0.05. Association analysis was conducted with the genome-wide efficient mixed-model association (GEMMA) software [66]. The Bonferroni-corrected significance threshold was applied to identify significantly associated variants, using the univariate linear mixed model, which is a general and widely used model with the following formula:

$$\mathbf{y} = \mathbf{W}\boldsymbol{\alpha} + \mathbf{x}\boldsymbol{\beta} + \mathbf{u} + \boldsymbol{\varepsilon};$$

$$\mathbf{u} \sim \text{MVN}_n(0, \lambda\tau^{-1}\mathbf{K}), \boldsymbol{\varepsilon} \sim \text{MVN}_n(0, \tau^{-1}\mathbf{I}_n)$$

where \mathbf{y} is an n -vector of quantitative traits (or binary disease labels) for n individuals; $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_c)$ is an $n \times c$ matrix of covariates (fixed effects) including a column of 1s (intercept); $\boldsymbol{\alpha}$ is a c -vector of the corresponding coefficients including the intercept; \mathbf{x} is an n -vector of marker genotypes; $\boldsymbol{\beta}$ is the effect size of the marker; \mathbf{u} is an n -vector of random effects; $\boldsymbol{\varepsilon}$ is an n -vector of errors; τ^{-1} is the variance of the residual errors; λ is the ratio between the two variance components; \mathbf{K} is a known $n \times n$ relatedness matrix; \mathbf{I}_n is an $n \times n$ identity matrix; and MVN_n denotes the n -dimensional multivariate normal distribution. We adjusted for potential population stratification using the first three principal components. To reduce the likelihood of false positive signals, we set the genome-wide significance threshold to 1.46×10^{-9} . This value corresponds to a P value of 0.05 divided by the total number of SNPs and INDELs, as determined by the Bonferroni correction.

Identification of selection signatures

We applied XP-EHH [67] and XP-CLR [68] analyses and used data from 145 Tibetan sheep (altitude > 3000 m a.s.l.) and 119 low-altitude sheep (altitude < 1200 m a.s.l.). These estimates were determined using 50-kb window that slides in 10-kb increments across the genomes. The threshold value corresponds to the top 1% of the XP-EHH values ($\text{XP-EHH} < -1.74$ or $\text{XP-EHH} > 1.99$) and XP-CLR values ($-\log_{10} \text{XP-CLR} > 1.943$).

RNA sequencing

Total RNA of 12 organs (cerebrum, hypothalamus, lung, muscle, heart, liver, kidney, spleen, rumen, small intestine, large intestine, and skin) from 8 adult sheep ($n = 4$ Tibetan sheep raised at high altitudes and $n = 4$ Hu sheep raised at low altitudes) was extracted using the TRIzol reagent (Catalog No. 15596026CN, Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's instructions. The RNA samples were subjected to quality control using a NanoDrop 2000 spectrophotometer (Catalog No. ND2000, Thermo Fisher Scientific), agarose gel electrophoresis, and the Agilent 2100 Bioanalyzer (Catalog No. 2001, Agilent Technologies, Santa Clara, CA). The RNA sequencing (RNA-seq) library construction and paired-end sequencing were done at Novogene according to the company's pipeline. The average number of reads per sample was 30,195,473.

From the raw RNA-seq data, adapter sequences, reads with more than 5% unknown nucleotides, and reads with many low-quality bases (more than half of the bases' qualities were less than 10) were removed with Trimmomatic (v0.36) [60]. Transcript abundance was quantified in transcripts per million (TPM) using kallisto (v0.46.1; <https://pachterlab.github.io/kallisto>) [69] with default parameters and an index built from the

GCF_016772045.1_ARS-UL_Ramb_v2.0_rna.gz from Ensembl. Transcript abundance was aggregated to the gene level using the tximport package [70] of the R software. The edgeR package [71] was used to calculate the differential expression of genes with a minimum fold change of 1 and a false discovery rate (FDR) lower than 0.05. GO term annotations were performed using Metascape [72]. The association between messenger RNA (mRNA) expression of candidate genes and altitude (from 0 to 5000 m a.s.l.) was tested with the R lm() function. Gene sets with the top 100 genes and a *P* value less than 0.05 were selected for GO term analysis. We utilized all genes in the genome as the enrichment background. Terms that met the following criteria were gathered and clustered based on their membership similarities: a *P* value less than 0.01, a minimum count of 3, and an enrichment factor greater than 1.5 (the enrichment factor is the ratio between the observed counts and the counts expected by chance). *P* values were calculated using the cumulative hypergeometric distribution. Exon-specific mRNA expression was quantified using QTLtools [73].

Determination of SVs

SVs in candidate regions were investigated using the Integrative Genomics Viewer (IGV) in PacBio HiFi sequencing data [74]. The mosdepth software (v0.2.2) [75] was used to calculate the sequencing coverage within candidate regions. The coverage at the candidate region (Chr15:47,954,795–48,033,976 bp) was extracted and divided by the average coverage of chromosome 15 to differentiate between haplotypes A and B. The diplotypes were defined as follows: haplotype A frequency ≥ 0.75 , homozygous haplotype A; $0.25 \leq$ haplotype A frequency < 0.75 , heterozygous haplotype A; and $0 \leq$ haplotype A frequency < 0.25 , homozygous haplotype B. Pairwise linear alignments between the genomes of different breeds and species at different altitudes were performed using the minimap2 software [76]. The genomes of argali (JAKZEL000000000.1) [54], the high-altitude Qiaoke sheep (ASM2241668v1, GCA_022416685.1) [46], and the low-altitude Texel sheep (Oar_v4.0, GCA_000298735.2) [77] were used to investigate SVs.

Validation of SVs using PCR amplification

Genomic DNA was extracted from blood samples using a Blood Genomic DNA Mini Kit (Catalog No. CW2087S, CWBIO, Taizhou, China). Then, the extracted DNA was used for PCR-based genotyping. Six pairs of primers were designed to amplify regions located upstream, downstream, and within the deletion region. A thermostable Taq DNA polymerase (Catalog No. 10106ES03, Yeasen, Shanghai, China) was used for the PCR reaction, and the following PCR amplification program was used for all primer pairs: initial denaturation at 94°C for 3 min; 36 cycles of 94°C for 10 s, 58°C for 20 s, and 72°C for 1 min; and 72°C for 5 min. Amplified fragments were evaluated on a 1.5% agarose gel. Briefly, successful amplification of PCR products indicates no deletion, whereas failed amplification from the sample indicates deletion. Primer sequences used are listed in Table S7.

Blood parameter analysis

Fresh blood samples (5 ml) were collected from 10 adult Dorset sheep living at low altitudes and measured within 12 h of collection for their hematological profiles, including RBC count and Hb concentration, which were combined with previously published data of 215 adult sheep living at both high and low altitudes [44].

Functional analysis of total Hbs in blood

Blood samples ($\sim 200 \mu\text{l}$) were added to a 5 \times volume of ice-cold water and incubated on ice to lyse the RBCs for 30 min. The samples were then centrifuged at 20,000 *g* for 10 min to remove cell debris. Buffer was added to the supernatants to a final concentration of 0.01 M 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES)/0.2 M NaCl (pH 7.4), and passed through a PD-10 desalting column (Catalog No. 28-9493-23, GE Healthcare, Chicago, IL) equilibrated with 25 ml of 0.01 M HEPES (pH 7.4) to remove intracellular cofactors.

Hb solutions [0.1 mM Hb in 0.1 M HEPES/0.05 M ethylene diamine tetraacetic acid (EDTA) buffer (pH 7.4)] used for O₂ equilibrium measurement were freshly prepared in the absence (stripped) or presence of 0.1 M KCl and 0.2 mM DPG [78,79]. O₂ equilibrium curves were measured using a modified diffusion chamber method as previously described [80,81]. Curves were measured in 0.1 M HEPES buffer with pH 6.9 or pH 7.4 at 37°C to estimate the Bohr effect, and with pH 7.4 at 25°C to estimate the enthalpy of oxygenation.

Hill plots $\{\log_{10} [\text{fractional saturation}/(1 - \text{fractional saturation})] \text{ vs. } \log_{10} P_{\text{O}_2}\}$ constructed from these measurements were used to estimate the partial pressure of oxygen (P_{O_2}) at half saturation (P_{50}) and the cooperativity coefficient (n_{50}) from the χ -intercept and slope of these plots, respectively. O₂ equilibrium curves for Hb solutions were measured four times, whereas P_{50} was recoded as mean \pm standard error of mean (SEM) [24,51,82].

The Bohr factor was quantified by the slope of linear plots of $\log_{10} P_{50}$ as a function of pH in the range of pH 6.9 and pH 7.4 ($\Phi = \Delta \log_{10} P_{50}/\Delta \text{pH}$). The overall change of enthalpy for oxygenation at pH 7.4 was calculated by using the Van't Hoff Equation $\Delta H = 2.303R [\Delta \log_{10} P_{50}/(\Delta 1/T)]$, where *R* is the gas constant (8.314 J/mol/K) and *T* is the absolute temperature in Kelvin [83]. The $\Delta H'$ values were corrected by excluding the heat of O₂ in solution (-12.6 kJ/mol O_2) [84].

Ethical statement

The present study was approved by the Animal Care and Use Committee of Northwest A&F University (Approval No. 2020DK0802). All methods were carried out in accordance with relevant guidelines and regulations of Northwest A&F University, China.

Data availability

The raw sequence data reported in this study have been deposited in the Genome Sequence Archive [85] at the National Genomics Data Center [86], Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformatics (GSA: CRA011038), and are publicly accessible at <https://ngdc.cnbc.ac.cn/gsa>.

CRedit author statement

Chao Li: Data curation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing. **Bingchun Chen:** Investigation, Validation. **Suo Langda:** Resources, Validation. **Peng Pu:** Validation. **Xiaoja Zhu:** Investigation, Writing – review & editing. **Shiwei Zhou:** Resources. **Peter Kalds:** Writing – original draft, Writing – review & editing. **Ke Zhang:** Resources. **Meenu Bhati:** Formal

analysis. **Alexander Leonard:** Formal analysis. **Shuhong Huang:** Formal analysis. **Ran Li:** Formal analysis, Investigation. **Awang Cuoji:** Resources. **Xiran Wang:** Validation. **Haolin Zhu:** Validation. **Yujiang Wu:** Conceptualization, Resources, Funding acquisition. **Renqin Cuomu:** Resources. **Ba Gui:** Resources. **Ming Li:** Formal analysis, Investigation. **Yutao Wang:** Resources. **Yan Li:** Resources. **Wenwen Fang:** Formal analysis. **Ting Jia:** Resources. **Tianchun Pu:** Resources. **Xiangyu Pan:** Formal analysis. **Yudong Cai:** Formal analysis. **Chong He:** Formal analysis. **Liming Wang:** Resources. **Yu Jiang:** Investigation. **Jian-Lin Han:** Conceptualization, Writing – review & editing. **Yulin Chen:** Conceptualization, Investigation, Funding acquisition. **Ping Zhou:** Conceptualization, Resources. **Hubert Pausch:** Conceptualization, Data curation, Formal analysis, Investigation, Supervision, Writing – review & editing. **Xiaolong Wang:** Conceptualization, Data curation, Formal analysis, Investigation, Resources, Supervision, Writing – review & editing, Funding acquisition. All authors have read and approved the final manuscript.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (<https://doi.org/10.1093/gpbjnl/qzae030>).

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported by the Tibetan R&D Program, China (Grant No. XZ202101ZD0001N), the China Agriculture Research System (Grant No. CARS-39), and the National Natural Science Foundation of China (Grant Nos. 31900313, 32161143010, and 31972526). We would like to thank Naveen Kumar Kadri and Danang Crysnanto for their helpful discussions. We appreciate the Chinese Scholarship Council's support for Chao Li's stay at ETH Zürich, as well as the High-Performance Computing (HPC) platform of Northwest A&F University for providing compute resources.

ORCID

0000-0002-2769-7616 (Chao Li)
 0009-0008-0908-3056 (Bingchun Chen)
 0009-0008-9458-5906 (Suo Langda)
 0000-0001-6960-3836 (Peng Pu)
 0000-0001-5330-1278 (Xiaoqia Zhu)
 0009-0008-7732-9257 (Shiwei Zhou)
 0000-0002-4276-483X (Peter Kalds)
 0000-0002-2392-1545 (Ke Zhang)
 0000-0003-4955-2255 (Meenu Bhati)
 0000-0001-8425-5630 (Alexander Leonard)
 0000-0003-2302-4200 (Shuhong Huang)
 0000-0002-8584-4100 (Ran Li)
 0009-0000-6736-1327 (Awang Cuoji)
 0009-0008-4973-6963 (Xiran Wang)
 0009-0000-6210-9026 (Haolin Zhu)
 0000-0002-3915-4889 (Yujiang Wu)
 0009-0002-2848-7511 (Renqin Cuomu)

0000-0002-8127-4809 (Ba Gui)
 0000-0001-7049-8666 (Ming Li)
 0000-0003-0580-5820 (Yutao Wang)
 0000-0003-1446-7253 (Yan Li)
 0000-0003-2016-2544 (Wenwen Fang)
 0009-0005-8872-2650 (Ting Jia)
 0009-0000-2485-6792 (Tianchun Pu)
 0000-0001-6841-7652 (Xiangyu Pan)
 0000-0001-5902-0901 (Yudong Cai)
 0000-0001-5977-6629 (Chong He)
 0000-0001-7088-3193 (Liming Wang)
 0000-0003-4821-3585 (Yu Jiang)
 0000-0002-1527-3963 (Jian-Lin Han)
 0000-0001-5679-4055 (Yulin Chen)
 0009-0007-3442-1232 (Ping Zhou)
 0000-0002-0501-6760 (Hubert Pausch)
 0000-0003-1620-1344 (Xiaolong Wang)

References

- [1] Chen FH, Dong GH, Zhang DJ, Liu XY, Jia X, An CB, et al. Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 BP. *Science* 2015;347:248–50.
- [2] Meyer MC, Aldenderfer MS, Wang Z, Hoffmann DL, Dahl JA, Degering D, et al. Permanent human occupation of the central Tibetan Plateau in the early Holocene. *Science* 2017; 355:64–7.
- [3] Jeong C, Ozga AT, Witonsky DB, Malmstrom H, Edlund H, Hofman CA, et al. 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* 2016;113:7485–90.
- [4] Hu XJ, Yang J, Xie XL, Lv FH, Cao YH, Li WR, et al. The genome landscape of Tibetan sheep reveals adaptive introgression from argali and the history of early human settlements on the Qinghai-Tibetan Plateau. *Mol Biol Evol* 2019;36:283–303.
- [5] Ren L, Yang Y, Qiu M, Brunson K, Chen G, Dong G. 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* 2022;144:105627.
- [6] Chen SY, Duan ZY, Sha T, Xiangyu J, Wu SF, Zhang YP. Origin, genetic diversity, and population structure of Chinese domestic sheep. *Gene* 2006;376:216–23.
- [7] Liu J, Ding X, Zeng Y, Yue Y, Guo X, Guo T, et al. Genetic diversity and phylogenetic evolution of Tibetan sheep based on mtDNA D-loop sequences. *PLoS One* 2016;11:e0159308.
- [8] Du L, Li J, Ma N, Ma Y, Wang J, Yin C, et al. Animal genetic resources in China: sheep and goats. Beijing: China Agriculture Press; 2011.
- [9] Zhao P, Li S, He Z, Zhao F, Wang J, Liu X, et al. Physiology and proteomic basis of lung adaptation to high-altitude hypoxia in Tibetan sheep. *Animals (Basel)* 2022;12:2134.
- [10] Witt KE, Huerta-Sanchez E. Convergent evolution in human and domesticated adaptation to high-altitude environments. *Philos Trans R Soc Lond B Biol Sci* 2019;374:20180235.
- [11] Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, et al. Genetic evidence for high-altitude adaptation in Tibet. *Science* 2010;329:72–5.
- [12] Huerta-Sanchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 2014;512:194–7.
- [13] Ouzhuluobu, He Y, Lou H, Cui C, Deng L, Gao Y, et al. *De novo* assembly of a Tibetan genome and identification of novel structural variants associated with high-altitude adaptation. *Natl Sci Rev* 2020;7:391–402.
- [14] Quan C, Li Y, Liu X, Wang Y, Ping J, Lu Y, et al. Characterization of structural variation in Tibetans reveals new

- evidence of high-altitude adaptation and introgression. *Genome Biol* 2021;22:159.
- [15] Yu L, Wang GD, Ruan J, Chen YB, Yang CP, Cao X, et al. Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. *Nat Genet* 2016;48:947–52.
- [16] Chiou KL, Janiak MC, Schneider-Crease IA, Sen S, Ayele F, Chuma IS, et al. Genomic signatures of high-altitude adaptation and chromosomal polymorphism in geladas. *Nat Ecol Evol* 2022; 6:630–43.
- [17] Miao B, Wang Z, Li Y. Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the gray wolf from the Tibetan plateau. *Mol Biol Evol* 2017;34:734–43.
- [18] Wang MS, Wang S, Li Y, Jhala Y, Thakur M, Otecko NO, et al. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. *Mol Biol Evol* 2020; 37:2616–29.
- [19] Gou X, Wang Z, Li N, Qiu F, Xu Z, Yan D, et al. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res* 2014; 24:1308–15.
- [20] Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet* 2013;45:1431–8.
- [21] Newman JH, Holt TN, Cogan JD, Womack B, Phillips JA 3rd, Li C, et al. Increased prevalence of *EPAS1* variant in cattle with high-altitude pulmonary hypertension. *Nat Commun* 2015; 6:6863.
- [22] Wang MS, Li Y, Peng MS, Zhong L, Wang ZJ, Li QY, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Mol Biol Evol* 2015;32:1880–9.
- [23] Zhang Q, Gou W, Wang X, Zhang Y, Ma J, Zhang H, et al. Genome resequencing identifies unique adaptations of Tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments. *Genome Biol Evol* 2016;8:765–76.
- [24] Signore AV, Storz JF. Biochemical pedomorphosis and genetic assimilation in the hypoxia adaptation of Tibetan antelope. *Sci Adv* 2020;6:eabb5447.
- [25] Kumar C, Song S, Jiang L, He X, Zhao Q, Pu Y, et al. Sequence characterization of *DSG3* gene to know its role in high-altitude hypoxia adaptation in the Chinese cashmere goat. *Front Genet* 2018;9:553.
- [26] Song S, Yao N, Yang M, Liu X, Dong K, Zhao Q, et al. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). *BMC Genomics* 2016;17:122.
- [27] Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, et al. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci Rep* 2016;6:38932.
- [28] Li C, Wu Y, Chen B, Cai Y, Guo J, Leonard AS, et al. Markhor-derived introgression of a genomic region encompassing *PAPSS2* confers high-altitude adaptability in Tibetan goats. *Mol Biol Evol* 2022;39:msac253.
- [29] Wu D, Liu Y, Chen W, Shao J, Zhuoma P, Zhao D, et al. How placenta promotes the successful reproduction in high-altitude populations: a transcriptome comparison between adaptation and acclimatization. *Mol Biol Evol* 2022;39:msac120.
- [30] Yang Z, Bai C, Pu Y, Kong Q, Guo Y, Ouzhuluobu, et al. Genetic adaptation of skin pigmentation in highland Tibetans. *Proc Natl Acad Sci U S A* 2022;119:e2200421119.
- [31] Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, et al. The yak genome and adaptation to life at high altitude. *Nat Genet* 2012; 44:946–9.
- [32] Lv FH, Cao YH, Liu GJ, Luo LY, Lu R, Liu MJ, et al. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. *Mol Biol Evol* 2022;39:msab353.
- [33] Wu DD, Yang CP, Wang MS, Dong KZ, Yan DW, Hao ZQ, et al. Convergent genomic signatures of high-altitude adaptation among domestic mammals. *Natl Sci Rev* 2020;7:952–63.
- [34] Chen ZH, Xu YX, Xie XL, Wang DF, Aguilar-Gomez D, Liu GJ, et al. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Commun Biol* 2021;4:1307.
- [35] Li X, Yang J, Shen M, Xie XL, Liu GJ, Xu YX, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun* 2020;11:2815.
- [36] Naval-Sanchez M, Nguyen Q, McWilliam S, Porto-Neto LR, Tellam R, Vuocolo T, et al. Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nat Commun* 2018;9:859.
- [37] Pan Z, Li S, Liu Q, Wang Z, Zhou Z, Di R, et al. Whole-genome sequences of 89 Chinese sheep suggest role of *RXFP2* in the development of unique horn phenotype as response to semi-feralization. *Gigascience* 2018;7:giy019.
- [38] Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol Biol Evol* 2016; 33:2576–92.
- [39] Davenport KM, Bickhart DM, Worley K, Murali SC, Salavati M, Clark EL, et al. An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. *Gigascience* 2022;11:giab096.
- [40] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- [41] Jiang Y, Wang X, Kijas JW, Dalrymple BP. Beta-globin gene evolution in the ruminants: evidence for an ancient origin of sheep haplotype B. *Anim Genet* 2015;46:506–14.
- [42] Xiang K, Ouzhuluobu, Peng Y, Yang Z, Zhang X, Cui C, et al. Identification of a Tibetan-specific mutation in the hypoxic gene *EGLN1* and its contribution to high-altitude adaptation. *Mol Biol Evol* 2013;30:1889–98.
- [43] Bao G, Zhao F, Wang J, Liu X, Hu J, Shi B, et al. Characterization of the circRNA-miRNA-mRNA network to reveal the potential functional ceRNAs associated with dynamic changes in the meat quality of the longissimus thoracis muscle in Tibetan sheep at different growth stages. *Front Vet Sci* 2022; 9:803758.
- [44] Li X, Yang S, Ma L, Yang J, Kong X, Han X, et al. Study on the blood physiological properties of hypoxia adaptation in Tibetan sheep. *Heilongjiang Anim Sci Vet Med* 2016;01:12–5. (in Chinese with an English abstract)
- [45] Wu Y, Tang Z, Yi X, Suo L, Zhu D, Song T. Assessment of blood biochemical indices in Northwest Tibetan cashmere goats. *China Herbiv Sci* 2016;36:65–7.
- [46] Li R, Gong M, Zhang X, Wang F, Liu Z, Zhang L, et al. A sheep pangenome reveals the spectrum of structural variations and their effects on tail phenotypes. *Genome Res* 2023;33:463–77.
- [47] Davenport KM, Massa AT, Bhattarai S, McKay SD, Mousel MR, Herndon MK, et al. Characterizing genetic regulatory elements in ovine tissues. *Front Genet* 2021;12:628849.
- [48] Li M, Sun C, Xu N, Bian P, Tian X, Wang X, et al. *De novo* assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on microchromosomes and subtelomeric regions. *Mol Biol Evol* 2022;39:msac066.
- [49] Kim J, Lee C, Ko BJ, Yoo DA, Won S, Phillippy AM, et al. False gene and chromosome losses in genome assemblies caused by GC content variation and repeats. *Genome Biol* 2022;23:204.
- [50] vonHoldt B, Fan Z, Ortega-Del Vecchyo D, Wayne RK. *EPAS1* variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. *PeerJ* 2017;5:e3522.
- [51] Signore AV, Yang YZ, Yang QY, Qin G, Moriyama H, Ge RL, et al. Adaptive changes in hemoglobin function in high-altitude

- Tibetan canids were derived via gene conversion and introgression. *Mol Biol Evol* 2019;36:2227–37.
- [52] Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun* 2018;9:2337.
- [53] Wu DD, Ding XD, Wang S, Wojcik JM, Zhang Y, Tokarska M, et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat Ecol Evol* 2018; 2:1139–45.
- [54] Li X, He SG, Li WR, Luo LY, Yan Z, Mo DX, et al. Genomic analyses of wild argali, domestic sheep, and their hybrids provide insights into chromosome evolution, phenotypic variation, and germplasm innovation. *Genome Res* 2022;32:1669–84.
- [55] Wailoo K. Sickle cell disease — a history of progress and peril. *N Engl J Med* 2017;376:805–7.
- [56] Blunt MH. *The blood of sheep: composition and function*. Berlin: Springer; 1975.
- [57] Wilson JB, Edwards WC, McDaniel M, Dobbs MM, Huisman TH. The structure of sheep hemoglobins: II. the amino acid composition of the tryptic peptides of the non- α chains of hemoglobins A, B, C, and F. *Arch Biochem Biophys* 1966;115:385–400.
- [58] Graham AM, McCracken KG. Convergent evolution on the hypoxia-inducible factor (HIF) pathway genes *EGLN1* and *EPAS1* in high-altitude ducks. *Heredity (Edinb)* 2019;122:819–32.
- [59] Jaakkola P, Mole DR, Tian YM, Wilson MI, Gielbert J, Gaskell SJ, et al. Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* 2001;292:468–72.
- [60] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.
- [61] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013;1303.3997.
- [62] Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 2016;32:292–4.
- [63] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81:559–75.
- [64] Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2:e190.
- [65] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10:giab008.
- [66] Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012;44:821–4.
- [67] Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007;449:913–8.
- [68] Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res* 2010;20:393–402.
- [69] Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016;34:525–7.
- [70] Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 2015;4:1521.
- [71] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
- [72] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10:1523.
- [73] Fort A, Panousis NI, Garieri M, Antonarakis SE, Lappalainen T, Dermizakis ET, et al. MBV: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics* 2017;33:1895–7.
- [74] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.
- [75] Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 2018;34:867–8.
- [76] Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 2021;37:4572–4.
- [77] Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 2014;344:1168–73.
- [78] Mairbäurl H, Weber RE. Oxygen transport by hemoglobin. *Compr Physiol* 2011;2:1463–89.
- [79] Pu P, Lu S, Niu Z, Zhang T, Zhao Y, Yang X, et al. Oxygenation properties and underlying molecular mechanisms of hemoglobins in plateau zokor (*Eospalax baileyi*). *Am J Physiol Regul Integr Comp Physiol* 2019;317:R696–708.
- [80] Weber RE. Use of ionic and zwitterionic (Tris/BisTris and HEPES) buffers in studies on hemoglobin function. *J Appl Physiol* 1992;72:1611–5.
- [81] Weber RE. Cationic control of O₂ affinity in lugworm erythrocytes. *Nature* 1981;292:386–7.
- [82] Zhu X, Guan Y, Signore AV, Natarajan C, DuBay SG, Cheng Y, et al. Divergent and parallel routes of biochemical adaptation in high-altitude passerine birds from the Qinghai-Tibet Plateau. *Proc Natl Acad Sci U S A* 2018;115:1865–70.
- [83] Jensen B, Storz JF, Fago A. Bohr effect and temperature sensitivity of hemoglobins from highland and lowland deer mice. *Comp Biochem Physiol A Mol Integr Physiol* 2016;195:10–4.
- [84] Antonini E, Brunoni M. Hemoglobin and myoglobin in their reactions with ligands. *Front Biol* 1971;21:27–31.
- [85] Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang K, et al. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics* 2021;19:578–83.
- [86] CNGB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformatics in 2022. *Nucleic Acids Res* 2022;50:D27–38.