

# Selection and introgression facilitated the adaptation of Chinese native endangered cattle in extreme environments

Xinfeng Liu<sup>1,2</sup> | Zhaohong Li<sup>2</sup> | Yubin Yan<sup>2</sup> | Ye Li<sup>2</sup> | Hui Wu<sup>1,2</sup> | Jie Pei<sup>3</sup> | Ping Yan<sup>3</sup> | Ruolin Yang<sup>2</sup> | Xian Guo<sup>3</sup> | Xianyong Lan<sup>1</sup> 

<sup>1</sup>College of Animal Science and Technology, Northwest A&F University, Yangling, China

<sup>2</sup>College of Life Sciences, Northwest A&F University, Yangling, China

<sup>3</sup>Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou, China

## Correspondence

Xian Guo, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences. No.335 Jianguoyuan, Lanzhou, Gansu 730050, China. Email: guoxian@caas.cn

Xianyong Lan, College of Animals Science and Technology, Northwest A&F University. No. 22 Xinong Road. Yangling, Shaanxi 712100, China. Email: lanxianyong79@nwsuaf.edu.cn

Ruolin Yang, College of Life Sciences, Northwest A&F University. No.22 Xinong Road. Yangling, Shaanxi 712100, China. Email: desert.ruolin@gmail.com

## Funding information

China Agriculture Research System, Grant/Award Number: CARS-37; Hundred Talents Program of Shaanxi Province of China, Grant/Award Number: A289021612; National Natural and Science Foundation of China, Grant/Award Number: 31872331

[Corrections added on 20 December 2020, after first online publication: Correspondence section has been modified.]

## Abstract

Although persistent efforts have identified and characterized a few candidate genes and related biological processes with potential functions in the adaptation of many species to extreme environments, few works have been conducted to determine the genomic basis of adaptation in endangered livestock breeds that have been living in extreme conditions for more than thousands of years. To fill this gap, we sequenced the whole genomes of nine individuals from three Chinese native endangered cattle breeds that are living in high-altitude or arid environments. Phylogenetic and evolutionary history analyses of these three and other six breeds showed that the genetic structure of the cattle populations is primarily related to geographic location. Interestingly, we identified pervasive introgression from the yak to Zhangmu cattle (ZMC) that cover several genes (e.g., *NOS2*, *EGLN1* and *EPAS1*) involved in the hypoxia response and previously identified as positive selection genes in other species, which suggested that the adaptive introgression from yak may have contributed to the adaptation of ZMC to high-altitude environments. In addition, by contrasting the breeds in opposite living conditions, we revealed a set of candidate genes with various functions from hypoxia response, water metabolism, immune response and body shape change to embryo development and skeletal system development, etc., that may be related to high-altitude or arid adaptation. Our research provides new insights into the recovery and adaptation of endangered native cattle and other species in extreme environments and valuable resources for future research on animal husbandry to cope with climate change.

## KEYWORDS

adaptation, cattle, endangered, extreme environment, introgression, SNPs

Xinfeng Liu, Zhaohong Li, Ruolin Yang, Xian Guo, and Xianyong Lan are equally contributed.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

## 1 | INTRODUCTION

Cattle have been regarded as one of the most important animals domesticated by humans since the Neolithic period because they provide meat, milk, leather and traction for farming and transportation (Zhang et al., 2013). Based on genetic characteristics and phenotypic differences, most modern cattle breeds are divided into two groups, humpless taurine (*Bos taurus*) and humped indicine (*Bos indicus*) cattle, which were domesticated in the Fertile Crescent and the Indus Valley, respectively (Bickhart et al., 2016; Hiendleder et al., 2008). There are 53 native cattle breeds in China, which are further classified into three groups by their geographical locations, morphologies and sex chromosome polymorphisms: (a) southern cattle breeds largely descend from the indicine lineage, (b) northern cattle breeds belong to the taurine lineage and (c) cattle breeds from central China originated from taurine × indicine cattle hybrids (Lei et al., 2006; Zhang, 2011). The diverse geographic conditions and climates in China have resulted in some of the domestic breeds of cattle adapting to different extreme environments during evolution. For example, Zhangmu cattle (ZMC) has long been distributed in Nyalam County, Gyirong County and Tingri County of Tibet, China, and is one of a few domestic cattle breeds living at an altitude of about 5,000 m in the world (Zhang, 2011). ZMC was suspected to be a hybrid breed formed from ancient hybridization of taurine cattle, indicine cattle and yak in natural conditions (Zhang, 2011). Anxi cattle (AXC), named after the former Anxi County in Gansu Province, is a local breed endemic to the Gobi Desert of the “ancient Silk Road-Hexi Corridor.” Due to the extremely harsh environment, AXC is thought to be the only local cattle breed living in this area (Zhang, 2011). Qaidam cattle (QDMC), another local taurine breed, which was domesticated from ancient Mongolian cattle with a strong resistance to alpine climates and mosquitos, is mainly distributed on the edge of Qaidam Basin (approximately 3,000 m in altitude) in Qinghai Province. QDMC has adapted to swamps and becomes a valuable genetic resource in this region (Zhang, 2011). However, these breeds are threatened by rapid climate changes, such as global warming in recent years (Easterling et al., 2000). In addition, increased human activities are another factor contributing to the decline in the species (Chen et al., 2019). It was estimated that the numbers of ZMC and AXC are less than 300 and thus on the verge of extinction. The number of QDMCs is less than 9,000 and is endangered (Zhang, 2011).

In recent years, genome-wide selection analysis has been performed to identify adaptive genetic variations in a variety of domesticated animals living in extreme environments, resulting in a lot of important findings. Tibetan cattle has received the yak alleles of *EGLN1*, *EGLN2* and *HIF3 $\alpha$*  in the HIF pathway by introgression (Wu et al., 2018). The genes involved in mitochondrial oxidative phosphorylation and hypoxic response have played a crucial role in the adaptation of yak to the extreme environment of the plateau (Qiu et al., 2012). *SOCS2* and *GPX3* were shown under strong positive selection in Tibetan sheep and Taklimakan desert sheep, respectively (Yang et al., 2016). More recently, a systematic investigation on the genetic mechanism underlying the adaptation of domestic animals

(including Tibetan mastiff, Tibetan pig, Tibetan sheep, Tibetan goat, Tibetan horse, Tibetan cattle) to the Qinghai Tibet Plateau revealed that the convergent evolution of some genes, such as *EPAS1*, might have played an important role for human and many domestic animals adapting to plateau environment (Wu et al., 2018). These studies have led to a better understanding of the genetic mechanisms underlying species adaptation to extreme environments. However, to our knowledge, little research has characterized the genetic adaptations of endangered cattle to extreme environments based on whole-genome deep sequencing.

In this study, for the first time, we used deep whole-genome resequencing data to explore adaptive genetic mechanisms in three endangered cattle breeds living in either extremely high-altitude or arid environments. These studies are of great significance for the conservation and restoration of endangered livestock genetic resources.

## 2 | MATERIALS AND METHODS

### 2.1 | Ethics statement

According to “Guidelines on Ethical Treatment of Experimental Animals (2006) No. 398,” the sampling procedures were in compliance with the Ministry of Science and Technology, China. All procedures conducted in this study were approved by the Northwest A&F University Experimental Animal Manage Committee.

### 2.2 | Sample collection and sequencing

A total of nine samples from Zhangmu cattle (ZMC,  $n = 3$ ), Qaidam cattle (QDMC,  $n = 3$ ) and Anxi cattle (AXC,  $n = 3$ ) were collected (Figure S1a; Table S1). The ear tissue of each individual was used to extract DNA. Using the Illumina standard protocol, genomic libraries with insert sizes of ~350 bp were constructed and sequenced by the Illumina HiSeq System (Illumina). In addition, the genome sequence data of 24 individuals were downloaded from the NCBI database, including Tibetan cattle (TC,  $n = 5$ ), Kazakh cattle (KZKC,  $n = 3$ ), Mongolian cattle (MGC,  $n = 3$ ), Wenling cattle (WLC,  $n = 2$ ), Wannan cattle (WNC,  $n = 2$ ), Leiqiong cattle (LQC,  $n = 3$ ) and six Yaks (Figure S1a; Table S1).

### 2.3 | Reads mapping and variation identification

All cleaned pair-end sequence reads were mapped against the reference bovine genome (UMD 3.1; Zimin et al., 2009) using Burrows-Wheeler Aligner (BWA 0.7.5a) software (Li & Durbin, 2009) with the parameters “BWA aln -o 1 -L -i 5 -e 4 -n 2.” Aligned sequences were converted to raw BAM files using SAMtools for sorting and duplicating (Li & Durbin, 2009). Next, Genome Analysis Toolkit (GATK, version 3.7) UnifiedGenotyper (McKenna et al., 2010) was used to

identify single nucleotide polymorphism (SNPs) with default settings. The raw SNPs were filtered by requiring a minimum coverage depth of 5 and a maximum of 100, a minimum RMS (root mean square) mapping quality score of 40 and no gap present within a 3-bp window. Finally, we only retained high-quality autosomal bi-allelic SNPs for subsequent analysis. SNPs variants were annotated using ANNOVAR (Wang et al., 2010).

## 2.4 | Linkage disequilibrium analysis

The values of identity by state (IBS) for all the samples were evaluated using PLINK v1.90 (Purcell et al., 2007). Genome-wide linkage disequilibrium (LD) of cattle was calculated as the mean-squared correlation coefficient ( $r^2$ ) values for pairwise markers using Haploview v4.2 software with the parameters “-maxdistance 200 -minGeno 0.5 -dprime -missingCutoff 1 -minGeno 0.6 -minMAF 0.05 -hwcutoff 0.001 -dprime” (Barrett et al., 2005).

## 2.5 | Population differentiation and genetic structure analysis

The phylogenetic relationship of cattle was inferred based on the high-quality autosomal bi-allelic SNPs using the neighbour-joining (NJ) method in PHYLIP (v3.695; <http://evolution.genetics.washington.edu/phylip.html>). In this analysis,  $p$ -distance was used and the produced phylogeny was visualized using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). The smartPCA program in EIGENSOFT (v7.2.10; Patterson et al., 2006) was used for principal component analysis (PCA) of the 27 individuals. ADMIXTURE was used to infer the population structure (Alexander et al., 2009). The cluster number  $K$  was set to 2 or 3, with 200 iterations for each run. The degree of genetic differentiation between any two groups of breeds was measured by  $F_{ST}$  (Weir & Cockerham, 1984).

## 2.6 | Inference of demographic history

Pairwise sequential Markovian coalescent (PSMC) model was used to infer the history of effective population size ( $N_e$ ) of the native cattle over the last one million years (Li & Durbin, 2011). The parameter was set to “-N30 -t15 -r5 -p '4 + 25\*2 + 4 + 6'.” In addition, the multiple sequential Markovian coalescent (MSMC) model (Schiffels & Durbin, 2014) was used to infer the divergence time between two breeds of interest (samples marked in Table S2). Before the MSMC analysis, all the individuals were phased using Beagle (version 3.3.2; Browning & Browning, 2011). We defined the estimated divergence time between a pair of breeds as the first time point at which the cross-coalescence rate was at 0.5. Both MSMC and PSMC used the same parameters generation time ( $g$ : 6 years) and average mutation rate ( $\mu$ :  $1.26 \times 10^{-8}$  per base per generation) as used in Chen et al. (2018).

## 2.7 | Detection of admixture

To infer the migration events of yak and nine native cattle breeds, a population-level phylogeny analysis was performed using TreeMix (v.1.12; Pickrell & Pritchard, 2012). The program inferred an ML tree for the nine native cattle breeds (27 individuals) and the yak as the out-group, and then, the residual matrix was used to identify pairs of populations that showed poor fits in the ML tree. From 0 to 5, migration events were gradually added to the ML tree until 99% of the variance between the breeds could be explained. The command parameter was “-bootstrap -k 10,000 -m migration events.” To further detect the evidence of admixture across populations, an ancestry graph and three-population ( $f_3$ ) tests implemented in TreeMix were utilized to examine the presence of admixture (Pickrell & Pritchard, 2012; Reich et al., 2009).

## 2.8 | Detection of pairwise identity by descent

To examine the genetic contribution from yak to ZMC, the genomic regions of the potential introgression between the yak and ZMC were detected by calculating the frequency of shared identical by descent (IBD) by referring to Bosse et al. (2014). Specifically, the autosomal sequences of all individuals were phased and IBD chunks were inferred using Beagle fastIBD (version 3.3.2) under default settings (Browning & Browning, 2011). Next, the IBD chunks between donator (yak,  $n = 6$ ) and recipient (ZMC,  $n = 3$ ) population or between ZMC and the remaining cattle population ( $n = 24$ ) were extracted to count the frequency of shared IBD with a window size of 20 kb and a step size of 10 kb. The relative frequency of IBD (rIBD) was then calculated as follows: rIBD value (yak and ZMC) – rIBD value (ZMC and remaining cattle breeds). Ultimately, genomic regions with negative rIBD values were identified as potential introgression from yak to ZMC. We further inspected a few potential introgression regions of interest with regard to  $F_{ST}$ ,  $\theta_{\pi}$ , Tajima's  $D$  (Tajima, 1989),  $D_{xy}$  (Nei, 1987) and the topology of induced NJ trees.

## 2.9 | Selective signals in extreme environments and nonextreme environments

For each of the nine cattle breeds, we first assigned it into two opposite groups (extreme environment groups vs. nonextreme environment groups) based on either the altitudes (altitude >3,000 m vs. <1,300 m) or the annual precipitation (average annual precipitation <350 mm vs. >500 mm) information (Table S13). Next, to detect the potential local selection regions of cattle genome corresponding to a certain extreme environment, we performed a sliding window analysis (with 50-kb windows and 10-kb steps) using three metrics:  $\theta_{\pi}$ ,  $F_{ST}$  and Tajima's  $D$ . An idea region with strong positive selection signal would be characteristic of high relative polymorphism level, high  $F_{ST}$  between groups and a lower negative Tajima's  $D$  in the extreme environment group than in the nonextreme environment group. The

average values of  $\theta_{\pi}$ ,  $F_{ST}$  and Tajima's  $D$  were calculated for each window using VCFtools (Danecek et al., 2011). It must be noted that apart from selection, demography can also influence Tajima's  $D$  value, so we should combine Tajima's  $D$  with other population genetic metrics (such as  $\theta_{\pi}$ ) to infer whether a local region is likely to be a target of positive selection (Tajima, 1989). We considered the windows with the top 5%  $F_{ST}$  and relative polymorphism levels simultaneously as the candidate regions subjected to positive selection. Finally, DAVID 6.7 (<http://david.abcc.ncifcrf.gov/>) was used for GO and KEGG analyses of overlapping genes in the candidate regions.

### 3 | RESULTS AND DISCUSSION

#### 3.1 | Whole-genome genetic variation

We collected three endangered cattle breeds from Tibet and north-west of China (Figure S1a; Table S1). Whole-genome resequencing generated a total of 7.57 billion paired-end reads with an insert size of approximately 350 bp on average. Alignment with the reference genome of *B. taurus* (UMD3.1) showed an average depth of 35.12x (Table S2). To place these cattle into a more detailed context, we analysed these genomic data with available resequencing data from 18 individuals representing six different breeds living in northern or southern China (Figure S1a; Table S1). We detected approximately 46 million SNPs among 27 individuals. After quality control, approximately 25 million high-quality SNPs for subsequent analyse, 87.61% of which were present in the dbSNP database (Table S3), indicating the high reliability of the called SNPs in this study. At the chromosome level, the longer the chromosome, the more the SNPs (Figure S2a; Table S4). At the breed level, LQC has the most total SNPs, followed by ZMC, WLC, WNC, TC, MGC, AXC, QDMC and KZKC. Obviously, the number of private SNPs of ZMC is significantly higher than other northern cattle breeds, indicating that ZMC may have higher genetic diversity (Figure S2b; Table S3). Approximately 9.4 million SNPs are shared between the three groups (Figure S2c). The genomic distribution of the high-quality SNPs indicated that a majority of them are located in intergenic regions but with only 0.68% in the exon regions (Table S5). For exonic SNPs, 64,679 SNPs and 104,080 SNPs were synonymous and nonsynonymous, respectively, which resulted in a nonsynonymous/synonymous ratio of 0.621 (Table S5).

#### 3.2 | Genomic variation and linkage disequilibrium

The genome-wide nucleotide diversity ( $\theta_{\pi}$ ) for the nine native cattle breeds was between  $1.54 \times 10^{-3}$  and  $2.37 \times 10^{-3}$ , which was comparable with the values from other mammals and humans (Figure S2d; Tables S6 and S7). Remarkably, ZMC showed a higher nucleotide diversity than the other five taurine cattle breeds (Figure S2d; Table S6). The genetic diversity was also reflected in the number of heterozygous SNPs, that is, although the number of heterozygous

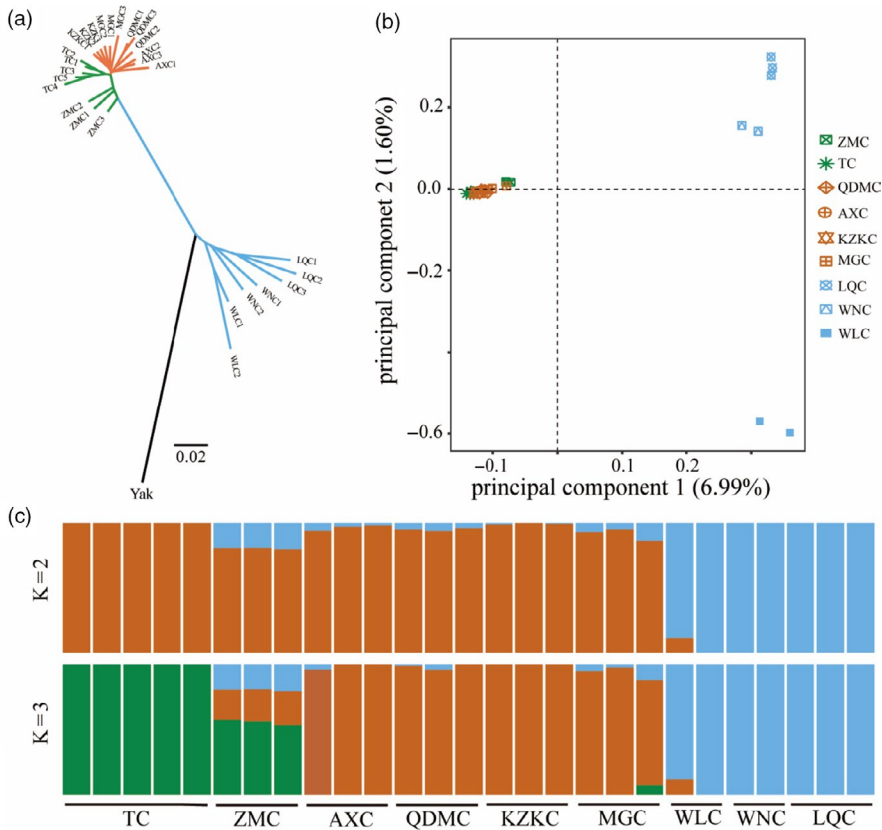
SNPs in ZMC is lower than that in the three southern indicine cattle breeds, the heterozygosity of ZMC is the highest among the analysed northern taurine cattle breeds (Table S3). Inbred individuals were not observed in the investigated cattle samples according to the IBS score ( $IBS < 0.9$ ; Table S8). LD analysis revealed that QDMC and AXC showed a slow decay rate and high level of LD, whereas the TC exhibited a rapid decay rate and a low level of LD (Figure S1b), which seems to be consistent with our finding that TC had the largest  $N_e$  among the cattle breeds to be analysed (see Section 3.4 for details).

#### 3.3 | Population genetic structure

To determine the phylogenetic relationship among the nine Chinese native breeds containing a total of 27 cattle individuals, we built a neighbour-joining (NJ) tree based on the whole-genome SNPs in autosomes using yak (*Bos grunniens*) as an out-group (Figure 1a). Notably, these native cattle samples were assigned to southern indicine cattle breeds (LQC, WNC and WLC) and northern taurine cattle breeds (TC, QDMC, MGC, KZKC and AXC) with a high level of the fixation index ( $F_{ST} = 0.260$ ), consistent with their ancestry and geographical distribution. Interestingly, ZMC is located between the two clades of indicine cattle and taurine cattle, and is closer to taurine cattle clade, suggesting that the ZMC has ancestor components of taurine and indicine (Figure 1a). The indicine cattle and taurine cattle were clearly separated by PCA (Figure 1b). We further performed ADMIXTURE analysis to explore the genetic composition of the samples with the number of clusters ( $K$ ) varying from 2 to 3 (Figure 1c). When  $K = 2$ , the samples were clearly divided into two groups, reflecting the divergence of taurine and indicine cattle, congruent with their geographic distribution and ancestral sources (Figure 1c). At  $K = 3$ , ZMC showed a clear admixture signature with genome ancestry with taurine and indicine cattle genetic background (Figure 1c). These results are largely consistent with the findings from the NJ tree and PCA and again suggested that ZMC possessed a unique evolutionary history.

#### 3.4 | Inference of demographic history

The population history for the 27 individuals was predicted using the PSMC model. In contrast, with the ancestral  $N_e$  of southern cattle breeds that might have mainly experienced a distinct decline during evolution, the North cattle breeds seemed to have undergone two distinct declines (Figure 2a and Figure S3). The first declines occurred ~0.9 Mya, coinciding with the Xixiabangma Glaciation (XG, 1.1–0.8 Mya; Zheng et al., 2002). After a long decline, the population of these taurine cattle breeds gradually recovered and ultimately peaked at ~40,000 years ago. Noticeably, although ZMC, such as other breeds of North ancestry, had a smaller  $N_e$  than that of the southern breeds, its  $N_e$  was significantly higher than the remainder of the northern taurine cattle during this peak period (Figure 2a and



**FIGURE 1** Population genetic analyses. (a) Neighbour-joining (NJ) phylogenetic tree of nine native cattle breeds (27 individuals) based on  $p$ -distances between individuals. The yak (*Bos grunniens*) was used as out-group. (b) Principal component analysis clustering pattern of individuals. Principal components 1 and 2 were used to represent the relationships among the 27 individuals. (c) ADMIXTURE result showing the genetic structure of Chinese native cattle. The length of each colour fragment represents the proportion of the individual genome inferred from the ancestral population ( $K = 2, 3$ )

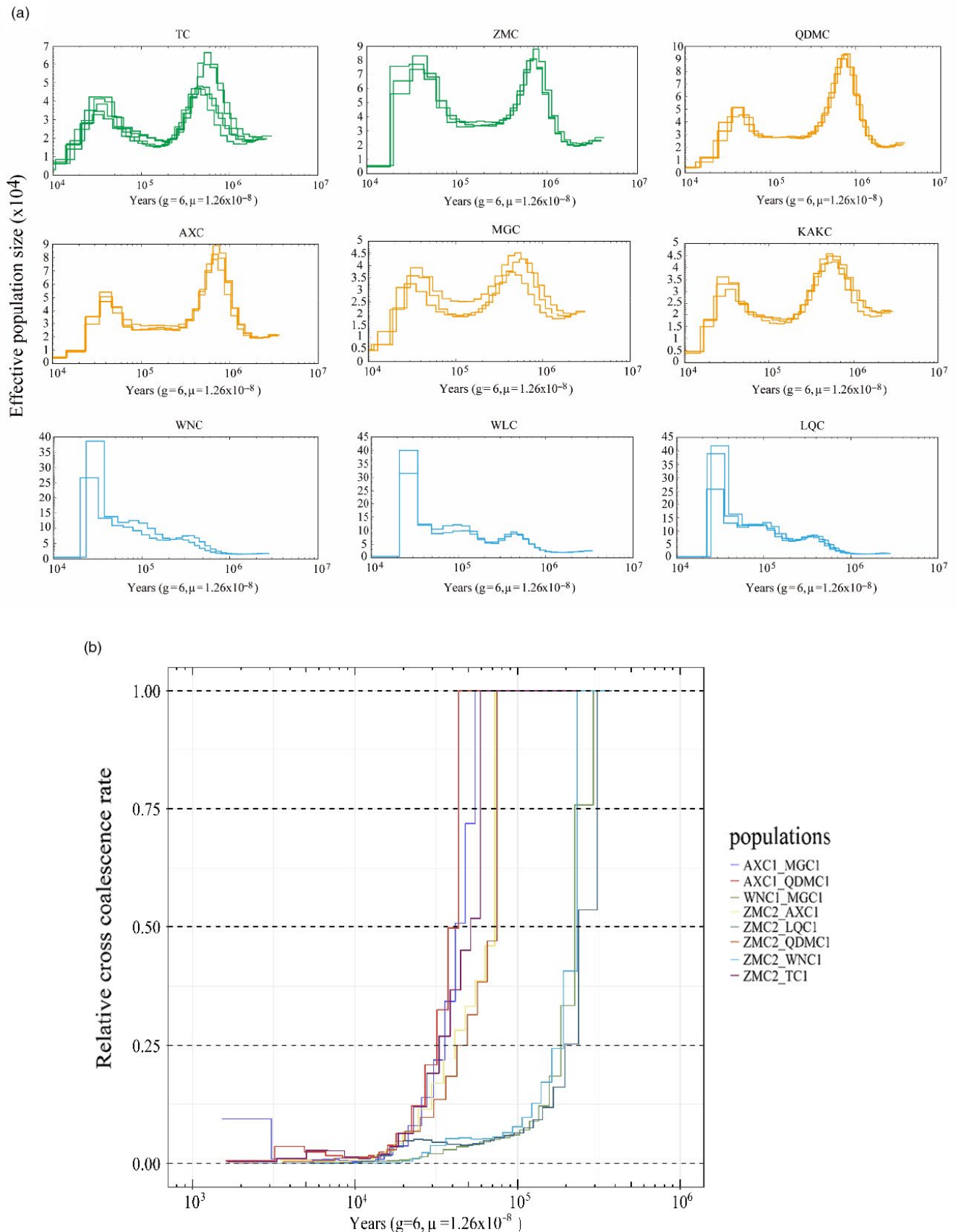
Figure S3). We speculate that this may be due to the mixture events between the ancestral ZMC and northern taurine cattle and southern indicine cattle, as reflected from the clustering pattern of the STRUCTURE analysis, resulting in a higher genetic diversity of ZMC than of other northern taurine cattle. Subsequently, the population size of all the breeds gradually decreased to a small number during the LGM (~20,000 years ago; Figure 2a and Figure S3; Lorenzen et al., 2011). Geological research shows that it was at the maximum value of the glacial ~20,000 years ago, and this extremely cold climate made it difficult for animals to survive and the  $N_e$  of animals declined sharply (Groenen et al., 2012). In addition, studies have also shown that during this period, human beings have expanded to Eurasia and America, occupying the living space of animals, resulting in the decline of the number of animals (Nielsen et al., 2017). We noticed that this period happens to coincide with the domestication of cattle. It was proposed that the domestication of an animal from their wild ancestor can cause an ancient bottleneck of these species (Wang et al., 2014). Therefore, we speculate that the domestication in itself may be another factor for the decline in the effective population size of livestock. Overall, consistent with the previous studies, our analysis provided an additional line of evidence for the changed effective population size in taurine cattle (Chen et al., 2018; Mei et al., 2017).

We used MSMC model to infer the divergence times between the three endangered breeds and other breeds inspected in this study (Figure 2b; Schiffels & Durbin, 2014). As the relative cross-coalescence rate (0.5) showed, ZMC may have diverged from the indicine

(LQC and WNC) cattle ~0.23 Mya during the Penultimate Glaciation (0.30–0.13 Mya; Figure 2b; Zheng et al., 2002). This inferred divergence time overlapped with a divergence time of taurine and indicine lineages (332–117 kya) estimated from the molecular genetic data (Figure 2b; Achilli et al., 2008; Bradley et al., 1996). Next, we found that the divergence time between ZMC and other northern cattle breeds was ~70–30 kya, which was obviously later than the divergence time between ZMC and southern cattle breeds (Figure 2b). The analysis results of PSMC and MSMC further supported the results of population genetic structure analysis.

### 3.5 | Inference of genetic admixture

The above analysis suggested that the increased genetic diversity in ZMC may have been due to an ancient gene flow between the ancestors of ZMC and the other cattle breeds. To further determine the signature of population admixture, TreeMix was used to evaluate possible admixture events among the nine local cattle breeds (Figure S4). An admixture graph obtained after adding five migration events optimally improved the model fit since it explained more than 99% of the genetic variance across populations (Figure S4). When  $M = 1-4$ , the obvious gene flow can be found from the northern taurine cattle to the ZMC (Figure S4). However, when  $M = 5$ , the method instead inferred weak gene flow from the ancestors of the southern indicine cattle to the ancestors of ZMC (Figure S4). Consequently, the TreeMix analysis suggested that the genome of extant ZMC might



**FIGURE 2** Demographic history of the native cattle. (a) History of effective population size of nine cattle breeds inferred from pairwise sequential Markov coalescence (PSMC). The PSMC model was used to infer the effective population size of the nine native cattle breeds over the past  $10^6$  years, the generation time ( $g = 6$ ) and the mutation rate ( $\mu = 1.26 \times 10^{-8}$ ), with one line per individual. The x-axis moves to the right to show the population size further back in time. (b) The relative cross-coalescence rates between the breeds were estimated using multiple sequential Markovian coalescent (MSMC), with four haplotypes per pair, generation time ( $g = 6$ ) and mutation rate ( $\mu = 1.26 \times 10^{-8}$ )

involve a mix of the genomic background of both indicine cattle and taurine cattle. To further evaluate the presence of admixture, we computed  $f_3$  statistics on all possible population triples (Population A; Population B; and Population C) using the TreeMix package. The extreme negative  $f_3$  statistic values indicate gene flow to Population A from both populations B and C. For MGC, this analysis produced six extreme negative Z-scores ( $-55.72$ ,  $-49.76$ ,  $-45.70$ ,  $-38.62$ ,  $-31.30$  and  $-30.89$ ) when using the populations of KZKC|LQC, WNC|KZKC, WLC|KZKC, TC|LQC, WNC|TC and WLC|TC as sources, respectively, which seemingly suggested MGC with a complex admixture history (Table S9), while we cannot rule out the possibility that the gene flow is from only one or two source breeds to MGC genome. In addition, using WNC and KZKC as sources produced Z-scores of  $-33.49$  for ZMC as an admixture population (Table S9), providing additional evidence of extensive gene flow during the evolution history of ZMC that contributed to the rich genetic diversity of this breed.

### 3.6 | Introgression mapping between yak and ZMC

Considering that both yak and ZMC are endemic to the Qinghai-Tibet Plateau, we examined possible yak introgression into ZMC. To this end, we calculated the rIBD statistics to infer the genomic regions of ZMC that might have been introgressed from yak. The results showed that there were  $\sim 48$  Mb ( $\sim 1.8\%$  of cattle genome) introgression regions ( $rIBD < 0$ ) between ZMC and yaks, covering 380 protein-coding genes (Figure 3a,b; Table S10). Noticeably, chromosome 3, chromosome 27 and chromosome 28 of ZMC contain several large introgression regions (Figure 3a and Figure S5), displaying a lower  $F_{ST}$  with yak than that in their adjacent nonintrogression regions (Figure S5). To consider the introgression process in the context of adaptive evolution, we performed enrichment analysis on the 380 genes overlapping with the introgression regions (Table S10). Accordingly, "olfactory transduction," "hematopoietic cell lineage," and "viral myocarditis" KEGG pathways, and "antigen processing and presentation," "MHC protein complex," "MHC class I protein complex" and "immune response" GO terms are the most significant pathways or GO terms with overrepresented candidate genes (Table S11). Interestingly, *NOS2*, *EPAS1* and *EGLN1* are annotated to "response to hypoxia" GO term (Table S11). *NOS2*,  $\sim 197$  kb in length, is overlapped with a continuous introgression region (Chr19:19860000–19960000) showing the lowest rIBD value (Figure 4a). Phylogeny analysis and *Dxy* assessment further corroborated this continuous region was likely to be introgressed from yak to ZMC (Figure 4a,d; Table S12). Moreover, compared with the adjacent nonintrogression regions, this continuous introgression region tended to exhibit a lower level of Tajima's  $D$  in ZMC population, a smaller  $F_{ST1}$  between ZMC and yak, but a larger  $F_{ST2}$  between ZMC and indicine cattle breeds (Figure 5a). Meanwhile, although the  $\theta_\pi$  displayed a larger fluctuation across windows probably due to the very limited samples, it seems that the nucleotide diversity in this introgression region was lower than that in the immediately flanking nonintrogression regions (Figure 5a). Together, this pattern probably suggested an episode of

ancient adaptive introgression. *NOS2* gene, encoding inducible nitric oxide synthase (iNOS), is responsible for the production of nitric oxide (NO) in organisms (Moncada & Higgs, 1991). Previous studies have shown that under hypoxic conditions, the expression of *NOS2* can be induced in many mammals and fish (Cameron et al., 2013; Kido et al., 2005; Tekin et al., 2010; Thompson et al., 2009; Yang et al., 2018). In addition, *NOS2* exhibits the strongest selection signal in the human high-altitude Andean population and is significantly associated with cardiovascular development and function (Crawford et al., 2017). Moreover, the *NOS2* gene was detected in Ladakhi cattle play a pivotal role in association with high-altitude adaptation through comparative transcriptome analysis (Verma et al., 2018). Incorporating these lines of evidence from our and previous studies, we cautiously speculate that the *NOS2* allele introgressed from yak may have driven the phenotypic evolution of ZMC.

Additionally, *EPAS1* and *EGLN1*, located within continuous 0.26 and 0.17 Mb introgression regions in chromosome 11 and chromosome 28, respectively, attracted our attention given their functions (Figure 4b,c). Both phylogenetic tree and *Dxy* analyses supported the two possible introgression events between yak and ZMC (Figure 4b,c,e,f; Table S12). *EPAS1* and *EGLN1* encode the hypoxia-inducible factor HIF-2 and prolyl 4-hydroxylase 2 (PHD2), respectively, both of which are located at the core of the hypoxia-inducible factor pathway and identified as key genes for hypoxia adaptation in Tibetan (Bigham & Lee, 2014; Lorenzo et al., 2014). Certain variants of *EGLN1* and *EPAS1* are associated with lower haemoglobin concentrations, which are beneficial for animals to avoid suffering from polycythemia (Xu et al., 2011). Obviously, the two introgression regions indeed showed a smaller Tajima's  $D$ , a lower level of differentiation between ZMC and yak but a higher level of differentiation between ZMC and indicine cattle breeds than their corresponding adjacent nonintrogression regions (Figure 5b,c). Again, the averaged  $\theta_\pi$  of the introgression regions seems to be lower, at least slightly, than that of their corresponding nonintrogression regions (Figure 5b,c), providing evidence for the positive selection of the two introgression regions. The yak has been living on the Qinghai-Tibet Plateau for millions of years and has evolved unique traits, such as the enlargement of the lungs and heart, to adapt to the extreme high-altitude environment (Chen et al., 2014). Domestic cattle were introduced onto the Qinghai-Tibet Plateau through human migration only thousands of years ago. The domestic cattle may be challenged by the high-altitude environment once they arrived at the Qinghai-Tibet Plateau. Luckily, the ancestors of domestic cattle obtained some foreign copies of homologous genes from yak; these genes, such as *EGLN1*, *EGLN2* and *HIF3a* from HIF pathways, have so important functions in the response to hypoxia that they may at least partially substitute the original alleles of specific cattle breeds, and eventually substantially contribute to their adaptation in the new habitat (Chen et al., 2018; Wu et al., 2018). Possibly in a similar way, the *EPAS1* has been found to be introgressed from Denisovans into Tibetans and from Tibetan wolves into the Tibetan mastiff, leading to the adaptive evolution of corresponding species (Huerta-Sánchez et al., 2014; Miao et al., 2017). Thus far, a few genes appear to be frequently adopted by different

species to conquer extreme habitats. Remarkably, introgression may provide an opportunity for those genes as a whole into the genomes of corresponding species, eventually contributing to the adaptive evolution of various species independently. In sum, our results suggested that the adaptive introgression from other species (e.g., yak) may have influenced the phenotypic evolution of domestic cattle.

It should be noted that we relied on phased genomic data to perform the introgression mapping analysis. Given the limited samples of ZMC, it would be hard to identify the rare and low-frequency haplotypes. Therefore, in this study, the detection of IBD that is based on the sharing of rare haplotype may be less accurate than it would be if using many more samples. To minimize this kind of risk that may lead to inflated results and spurious conclusions, we only retained two analyses whose input data are in the form of phased data, and all the other analyses were based on nonphased genomic data. Although the introgression analysis might not be completely confidential, these results at least provide a valuable resource of candidate genes for further experimental validation in the future, while we think that more samples are needed to draw a more reliable conclusion regarding the evolutionary history and selective forces of putative introgression of interest.

### 3.7 | Adaptive mechanisms in high-altitude environments

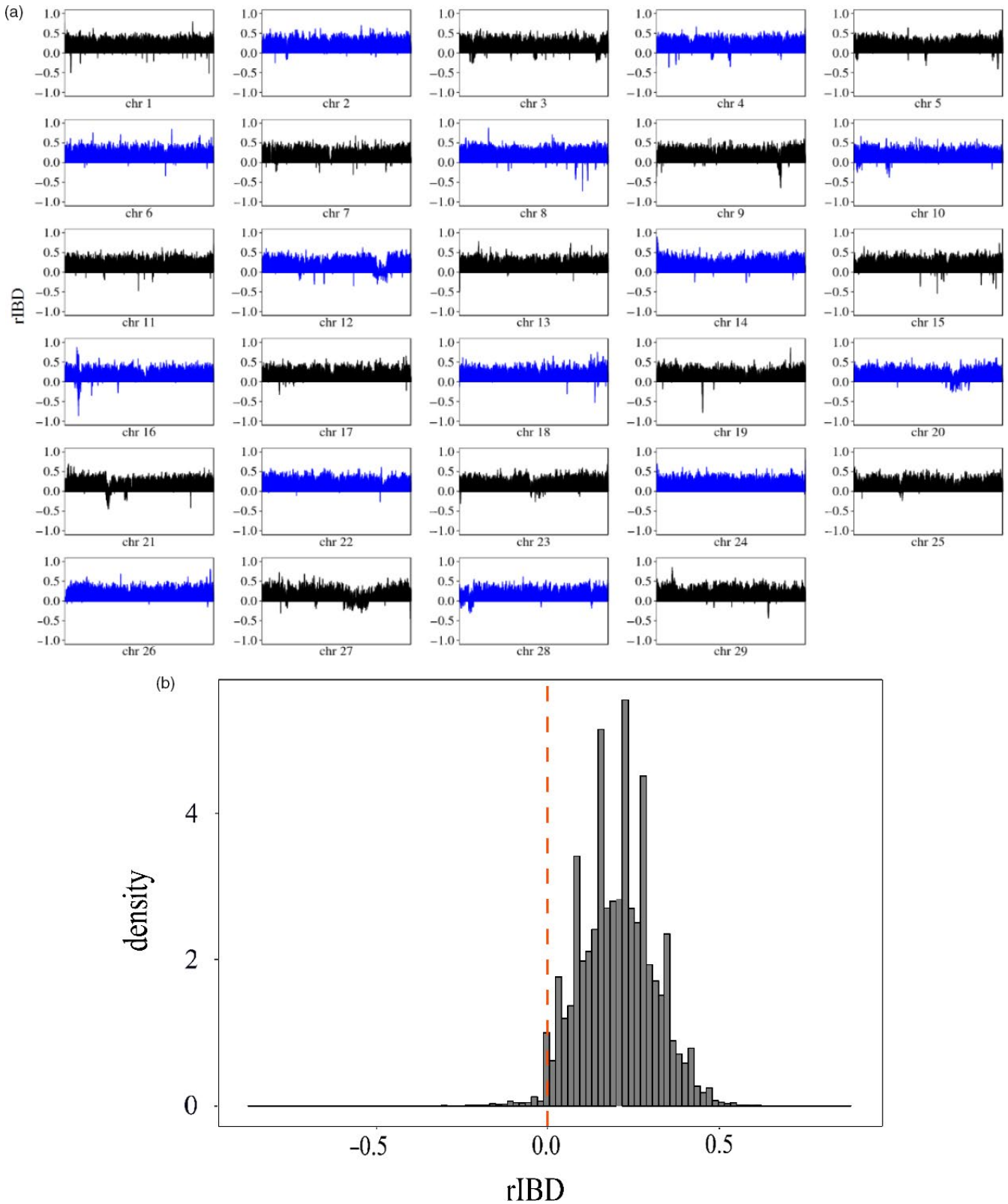
A more interesting question is to what extent the evolution of domestic cattle might be driven by selective forces related to environmental stress. To address this issue, we performed genome-wide selection scan analysis using two contrasting groups of cattle breeds that have lived in either high-altitude or low-altitude environments (Table S13). Using the top 5% of the  $F_{ST}$  values and  $\theta_{\pi}$  ratios as cut-offs, we obtained 1,056 candidate genes that might be related to high-altitude stress to varying degrees (Figure S6a; Table S14). By referencing the GO and KEGG pathway annotations, we found six genes in the classical HIF-1 (hypoxia-induced factors) pathway and 11 genes belonging to the GO term “the cellular-response/response to hypoxia” (Table S16), suggesting the potential roles of these genes to the hypoxia adaptation of high-altitude cattle breeds. Among the candidate genes, *PTPN9* attracted our attention, that is it showed higher  $F_{ST}$ ,  $\theta_{\pi}$  ratio, more negative Tajima's  $D$  than its adjacent nonintrogression regions in the high-altitude group, meaning a positive selection signature (Figure S6b). *PTPN9* is a member of protein tyrosine phosphatase (PTP) family with known functions in cell proliferation, differentiation and migration. *PTPN9* exerts its role through dephosphorylating protein substances, such as NSF, FOXO1, ERBB2, EGFR and VEGFR2, promoting homotypic vesicle fusion (Huynh et al., 2004), mediating insulin signalling in hepatocytes (Cho et al., 2006), suppressing breast cancer cell growth (Yuan et al., 2010) and regulating endothelial cell function (Hao et al., 2012). Besides, *PTPN9* has been identified as a positively selected gene in Tibetan sheep, possibly exerting its function via an interaction with *EGFR* in the classic HIF-1 pathway (Yang et al., 2016). Therefore, *PTPN9* is most likely to be an important

gene for the adaptation of domestic livestock to extreme high-altitude stress. In addition, seven genes in vascular smooth muscle contraction (VSMC) pathway and three genes in glycolysis/gluconeogenesis pathway (Table S16) were discovered. VSMC pathway functions in adjusting the diameter of blood vessels and the delivery of blood oxygen (Yang et al., 2016); glycolysis/gluconeogenesis pathway is particularly important for the energy metabolism of cattle under extreme hypoxic conditions. The recovery of these genes in this comparative analysis is consistent with a previous finding in the Tibetan antelope (Ge et al., 2013). The immune to response GO term also annotated some candidate genes, such as *IL6* gene (Table S16). *IL6* is an immune-related regulatory gene, encoding interleukin-6 cytokine. It can mediate the differentiation of macrophages in vivo, inhibit the production of inflammatory cytokines by macrophages and promote the survival and regeneration of damaged epithelium during inflammation (Grivennikov et al., 2009; Mauer et al., 2014). Therefore, the IL6-mediated immune response seems to play an important role in disease resistance and adaptation to harsh conditions for high-altitude cattle. Previous studies have revealed that genes involved in lipid metabolism play an important role in the adaptation of many high-altitude organisms to cold environments (Chen et al., 2016; Cheviron et al., 2012; Qiu et al., 2015). In this respect, we found two lipid metabolism-related genes, *B4GALNT1* and *PLIN2*, among the candidate genes (Table S16). Moreover, many genes with GO terms “embryo development” or “skeletal system development” were included in the candidate genes (Table S16). It has been reported that some genes with developmental functions undergo positive selection in other high-altitude species, such as the Tibetan human population (Xi et al., 2016) and Tibetan sheep (Yang et al., 2016). Considering the observation that the cattle population on the Qinghai-Tibet Plateau has a smaller body size and low energy consumption rate than low-altitude breeds, we speculate that the development regulatory network may be involved in the adaptive evolution of cattle at high-altitude regions by regulating the physical form of cattle body. Overall, our results provide many candidate genes that might relate to the high-altitude adaptation of cattle.

### 3.8 | Adaptive mechanisms in arid environments

Next, we investigated the genomic selection signal that might be associated with the adaptation of cattle in arid environment by comparing the cattle breeds lived in arid and humid environments (Table S13). Using the same method, 871 candidate genes were screened (Figure S7a; Table S15). Among them, 286 genes are also among the 1,056 candidate genes screened using altitude information, suggesting that these genes might be involved in adaptive responses to diverse environmental stress. Among the 871 genes screened in arid stress, four, three and ten genes, respectively, are in the arachidonic acid (AA) metabolism pathway, the renin-angiotensin system pathway and the oxytocin signalling pathway (Table S17). Previous researches indicated that these signalling pathways play significant

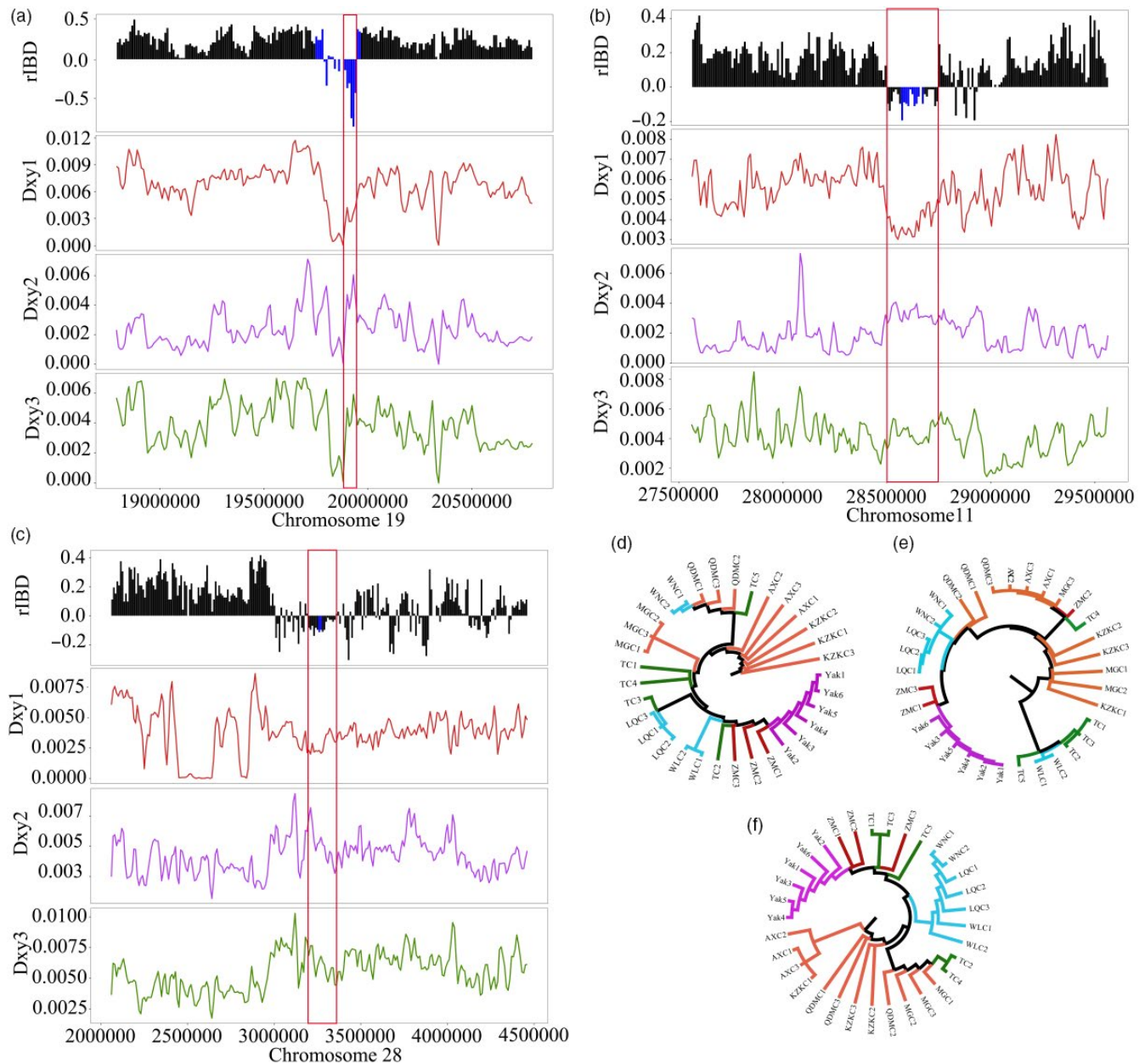




**FIGURE 3** Genome-wide pattern of introgression between the Zhangmu cattle (ZMC) and yak. (a) The x-axis represents the whole chromosome, and the y-axis represents the rIBD value. Negative rIBD indicated that the signal of introgression came from the yak. (b) Distribution of rIBD in ZMC and the yak (−1 to 0) or control population (0 to 1)

roles in regulating water retention and reabsorption in renal kidney cells and blood vessels (Breyer & Breyer, 2000; Maguire et al., 1997). Therefore, our analysis suggested that the arid adaptation of species

may be achieved via a complex signalling network orchestrating downstream effector genes. To provide a concrete example, the gene *PLA2G4B*, a member in the AA metabolism pathway, displayed high

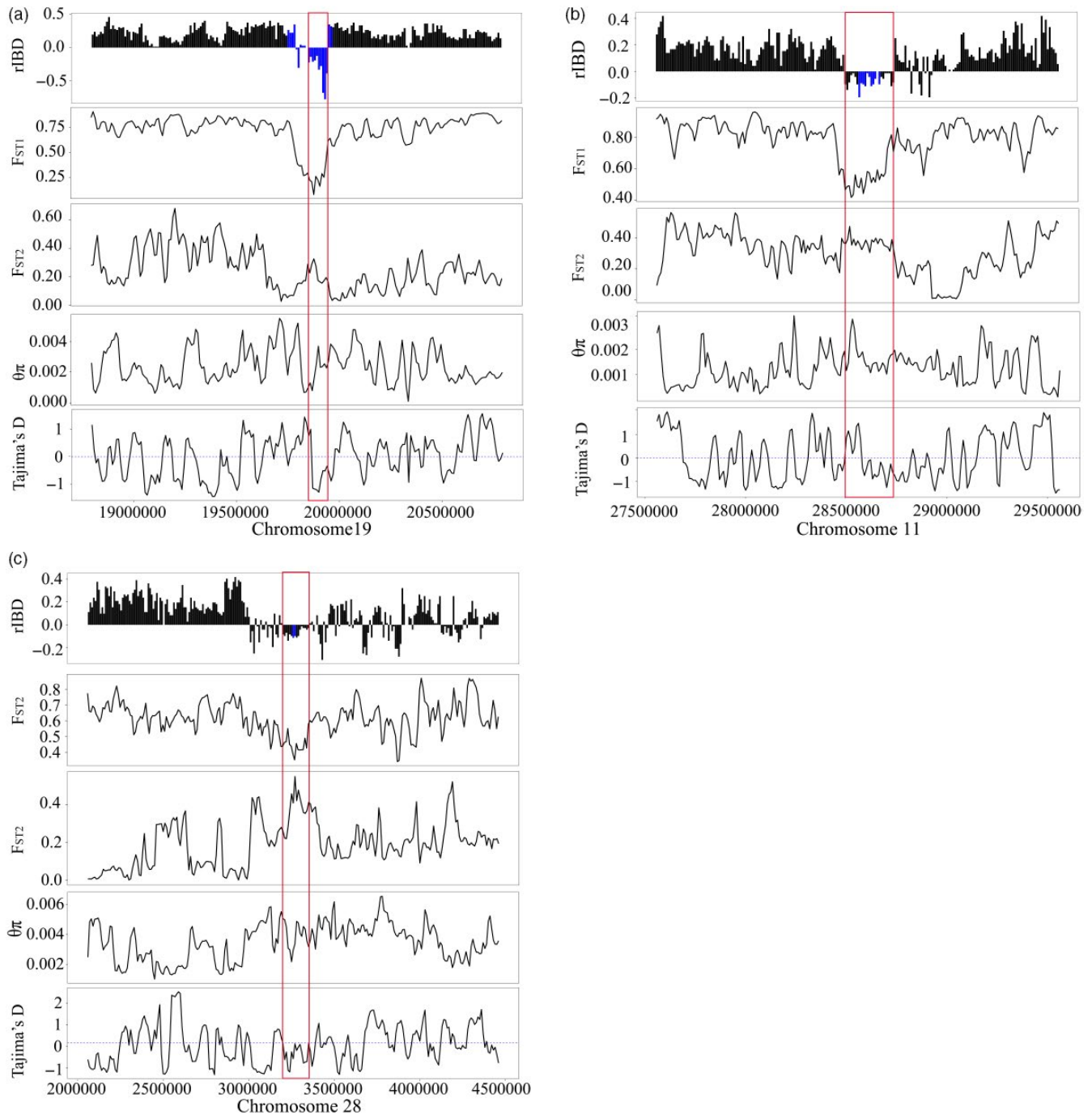


**FIGURE 4** Dxy and phylogeny analysis for adaptive introgression regions. (a–c) rIBD, Dxy1 (between Zhangmu cattle [ZMC] and yak), Dxy2 (between ZMC and taurine cattle) and Dxy3 (between ZMC and indicine cattle) values of the three continuous introgression regions in the chromosome 19 (19860000–19960000), chromosome 11 (28500000–28760000) and chromosome 28 (3200000–33700000), respectively. (d–f) Phylogenetic tree for the continuous introgression regions, respectively. The continuous introgression regions of the chromosome 19, chromosome 11 and chromosome 28 are marked by red boxes and the loci of *EPAS1*, *NOS2* and *EGLN1* are marked by blue bars

$F_{ST}$  and  $\theta_a$  ratio, and a lower Tajima's  $D$  value than the adjacent areas of this gene only for the arid group, indicating a signal of positive selection (Figure S7b). *PLA2G4B* belongs to the cytosolic phospholipase A2 protein family (Song et al., 1999). Previous studies have shown that genes located in the AA metabolism pathway play an important role in the conversion of AA to 19(S)-HETE (an effective renal vasodilator that stimulates water reabsorption; Bradley et al., 1996; Jirimutu et al., 2012). Hence, we speculate that *PLA2G4B* might have experienced adaptive evolution since the ancestors of cattle breeds entered into extreme arid conditions. This finding is consistent with the observation that the other three genes in the AA metabolic pathway are

under positive selection in Bactrian camels or sheep living in desert environments (Jirimutu et al., 2012; Yang et al., 2016).

In addition, six candidate genes belong to the pancreatic secretion pathway with the function of protein and carbohydrate digestion and absorption (Yang et al., 2016), which seems to well explain the characteristics of coarse feeding tolerance of livestock living in extremely arid conditions (Table S17). We also identified GO terms with diverse function annotations from immune response, lung development, skeletal system development, embryo development, response to xenobiotic stimulus and regulation of glucose metabolic process (Table S17), indicating the importance of body shape changes, energy optimization, disease defences



**FIGURE 5** Adaptive introgression signals from yak to Zhangmu cattle (ZMC). (a–c) Distribution patterns of rIBD,  $\theta_{\pi}$  (ZMC), Tajima's  $D$ ,  $F_{ST1}$  (between ZMC and yak) and  $F_{ST2}$  (between ZMC and indicine cattle) around the continuous introgression regions in the chromosome 19 (19860000–19960000), chromosome 11 (28500000–28760000) and chromosome 28 (3200000–33700000), respectively. The red boxes represent the continuous introgression regions, and blue bars represent the loci of *NOS2*, *EPAS1* and *EGLN1*

and stress responses to survival in arid environments. The function of these GO terms is highly consistent with the small size and high saline-alkali tolerance observed in sheep breeds in the Taklamakan Desert (Yang et al., 2016). Given the dry climate in north-west China with high soil salinity levels and a lack of forage resources, it is reasonable to speculate that these pathways should have played important roles in adapting cattle to extreme drought stress. In general, given the persistence of global warming and the increasing incidence of drought, these candidate genes

screened in this study would be a valuable resource for studying the adaptation of organisms to harsh environmental changes.

## 4 | CONCLUSION

In this study, for the first time, we used deep genome-wide polymorphism information on three native Chinese endangered cattle living in

extreme environments (i.e., high-altitude and arid areas in north-west China) to conduct a comprehensive analysis of their genetic backgrounds. We determined the phylogenetic relationship of the native endangered cattle and identified a complex gene admixture history between multiple breeds in the *Bos* subfamily. In particular, we found that ZMC may have obtained foreign alleles of at least three core genes (*NOS2*, *ENGL1* and *EPAS1*) of the “response to hypoxia” biological process from the yak by introgression, which is of great significance for the adaptation of ZMC to extremely high-altitude environments. Then, genome-wide comparison analyses between native cattle breeds from extreme and nonextreme environments revealed a variety of genes with diverse functions that may jointly contribute to the adaptation of cattle in plateau and arid environments. Overall, this work revealed the evolutionary history and adaptive characteristics of endangered cattle breeds living in extreme environments, which will provide useful guidance for the development of reasonable conservation plans in the future.

## ACKNOWLEDGEMENTS

This work was funded by the National Natural and Science Foundation of China (No. 31872331), China Agriculture Research System (No. CARS-37) and the Hundred Talents Program of Shaanxi Province of China (No. A289021612).

## CONFLICT OF INTEREST

None declared.

## AUTHOR CONTRIBUTIONS

X. Lan and X. G. conceived the research. X. Lan, R. Y. and Z. L. designed the research. X. G. prepared animal samples. X. Liu and Z. L. collected and analysed the data with inputs from all authors. X. Liu drafted the manuscript. X. Lan., R. Y., X. Liu., Z. L and Y. Y. revised the manuscript. X. Lan and R. Y. supervised the project. All authors read and approved the final manuscript.

## DATA AVAILABILITY STATEMENT

Our whole-genome Illumina sequencing reads have been deposited in the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) with the accession code (BioProject ID: PRJNA597241). The whole-genome sequence data of native cattle and Yaks are available at National Center for Biotechnology Information using Sequence Read Archive numbers SRR5507243, SRR5507244, SRR5507245, SRR6024571, SRR6024572, SRR5507258, SRR5507259, SRR5507260, SRR5507261, SRR5507262, SRR5507263, SRR6234762, SRR6234763, SRR6024573, SRR6024574, SRR5507188, SRR5507189, SRR5507190, SRR2059950, SRR2059951, SRR2059952, SRR2059962, SRR2059963 and SRR2059964.

## ORCID

Xianyong Lan  <https://orcid.org/0000-0003-2254-5805>

## REFERENCES

Achilli, A., Olivieri, A., Pellecchia, M., Ubaldi, C., Colli, L., Al-Zahery, N., Accetturo, M., Pala, M., Kashani, B. H., Perego, U. A., Battaglia,

- V., Fornarino, S., Kalamati, J., Houshmand, M., Negrini, R., Semino, O., Richards, M., Macaulay, V., Ferretti, L., ... Torroni, A. (2008). Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology*, 18(1), R157–R158. <https://doi.org/10.1016/j.cub.2008.01.019>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(6), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263–265. <https://doi.org/10.1093/bioinformatics/bth457>
- Bickhart, D. M., Xu, L., Hutchison, J. L., Cole, J. B., Null, D. J., Schroeder, S. G., Song, J., Garcia, J. F., Sonstegard, T. S., Van Tassell, C. P., Schnabel, R. D., Taylor, J. F., Lewin, H. A., & Liu, G. E. (2016). Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. *DNA Research*, 23(2), 253–262. <https://doi.org/10.1093/dnares/dsw013>
- Bigham, A. W., & Lee, F. S. (2014). Human high-altitude adaptation: Forward genetics meets the HIF pathway. *Genes & Development*, 28(20), 2189–2204. <https://doi.org/10.1101/gad.250167.114>
- Bosse, M., Megens, H.-J., Frantz, L. A. F., Madsen, O., Larson, G., Paudel, Y., Duijvesteijn, N., Harlizius, B., Hagemeijer, Y., Crooijmans, R. P. M. A., & Groenen, M. A. M. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nature Communication*, 5, 4392. <https://doi.org/10.1038/ncomms5392>
- Bradley, D. G., MacHugh, D. E., Cunningham, P., & Loftus, R. T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), 5131–5135. <https://doi.org/10.1073/pnas.93.10.5131>
- Breyer, M. D., & Breyer, R. M. (2000). Prostaglandin receptors: Their role in regulating renal function. *Current Opinion in Nephrology and Hypertension*, 9(1), 23–29. <https://doi.org/10.1097/00041552-200001000-00005>
- Browning, S. R., & Browning, B. L. (2011). A fast, powerful method for detecting identity by descent. *American Journal of Human Genetics*, 88(2), 173–182. <https://doi.org/10.1016/j.ajhg.2011.01.010>
- Cameron, J. S., DeWitt, J. P., Ngo, T. T., Yajnik, T., Chan, S., Chung, E., & Kang, E. (2013). Cardiac KATP channel alterations associated with acclimation to hypoxia in goldfish (*Carassius auratus* L). *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 164(4), 554–564. <https://doi.org/10.1016/j.cbpa.2012.12.020>
- Chen, C., Liu, Z., Pan, Q., Chen, X., Wang, H., Guo, H., & Shi, W. (2016). Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinixinyuan* n. ssp. *Molecular Biology and Evolution*, 33(5), 1337–1348. <https://doi.org/10.1093/molbev/msw017>
- Chen, F. H., Dong, G. H., Zhang, D. J., Liu, X. Y., Jia, X., An, C. B., Ma, M. M., Xie, Y. W., Barton, L., Ren, X. Y., Zhao, Z. J., Wu, X. H., & Jones, M. K. (2014). Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P. *Science*, 347(6219), 248–250. <https://doi.org/10.1126/science.1259172>
- Chen, L., Qiu, Q., Jiang, Y. U., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., Wang, J., Nie, W., Su, W., Liu, G., Li, Q., Fu, W., Pan, X., Liu, C., Yang, J., Zhang, C., Yin, Y., ... Wang, W. (2019). Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, 364, pii: eaav6202. <https://doi.org/10.1126/science.aav6202>
- Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., Hu, S., Huang, S., Zhang, H., Zheng, Z., Song, W., Ma, Z., Ma, Y., Dang, R., Zhang, Z., Xu, L., Jia, Y., Liu, S., Yue, X., ... Lei, C. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communication*, 9(1), 2337. <https://doi.org/10.1038/s41467-018-04737-0>

- Cheviron, Z. A., Bachman, G. C., Connaty, A. D., McClell, G. B., & Storz, J. F. (2012). Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22), 8635–8640. <https://doi.org/10.1073/pnas.1120523109>
- Cho, C. Y., Koo, S.-H., Wang, Y., Callaway, S., Hedrick, S., Mak, P. A., Orth, A. P., Peters, E. C., Saez, E., Montminy, M., Schultz, P. G., & Chanda, S. K. (2006). Identification of the tyrosine phosphatase PTP-MEG2 as an antagonist of hepatic insulin signaling. *Cell Metabolism*, 3(5), 367–378. <https://doi.org/10.1016/j.cmet.2006.03.006>
- Crawford, J. E., Amaru, R., Song, J., Julian, C. G., Racimo, F., Cheng, J. Y., Guo, X., Yao, J., Ambale-Venkatesh, B., Lima, J. A., Rotter, J. I., Stehlik, J., Moore, L. G., Prchal, J. T., & Nielsen, R. (2017). Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. *The American Journal of Human Genetics*, 101(5), 752–767. <https://doi.org/10.1016/j.ajhg.2017.09.023>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., & Mearns, L. O. (2000). Climate extremes: Observations, modeling, and impacts. *Science*, 289(5487), 2068–2074. <https://doi.org/10.1126/science.289.5487.2068>
- Ge, R.-L., Cai, Q., Shen, Y.-Y., San, A., Ma, L., Zhang, Y., Yi, X., Chen, Y., Yang, L., Huang, Y., He, R., Hui, Y., Hao, M., Li, Y., Wang, B. O., Ou, X., Xu, J., Zhang, Y., Wu, K., ... Wang, J. (2013). Draft genome sequence of the Tibetan antelope. *Nature Communication*, 4, 1858. <https://doi.org/10.1038/ncomms2860>
- Grivennikov, S., Karin, E., Terzic, J., Mucida, D., Yu, G. Y., Vallabhapurapu, S., Scheller, J., Rose-John, S., Cheroutre, H., Eckmann, L., & Karin, M. (2009). IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. *Cancer Cell*, 15(2), 103–113. <https://doi.org/10.1016/j.ccr.2009.01.001>
- Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H. J., Li, S., Larkin, D. M., Kim, H., Frantz, L. A., Caccamo, M., Ahn, H., Aken, B. L., Anselmo, A., Anthon, C., ... Schook, L. B. (2012). Analysis of pig genomes provide insight into porcine demography and evolution. *Nature*, 491(7424), 393–398. <https://doi.org/10.1038/nature11622>
- Hao, Q., Samten, B., Ji, H. L., Zhao, Z. J., & Tang, H. (2012). Tyrosine phosphatase PTP-MEG2 negatively regulates vascular endothelial growth factor receptor signaling and function in endothelial cells. *American Journal of Physiology-Cell Physiology*, 303(5), C548–C553. <https://doi.org/10.1152/ajpcell.00415.2011>
- Hiendleder, S., Lewalski, H., & Janke, A. (2008). Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. *Cytogenetic and Genome Research*, 120(1–2), 150–156. <https://doi.org/10.1159/000118756>
- Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y. U., Yi, X., He, M., Somel, M., Ni, P., Wang, B. O., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y. E., ... Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, 12(7513), 194–197. <https://doi.org/10.1038/nature13408>
- Huynh, H., Bottini, N., Williams, S., Cherepanov, V., Musumeci, L., Saito, K., Bruckner, S., Vachon, E., Wang, X., Kruger, J., Chow, C.-W., Pellecchia, M., Monosov, E., Greer, P. A., Trimble, W., Downey, G. P., & Mustelin, T. (2004). Control of vesicle fusion by a tyrosine phosphatase. *Nature Cell Biology*, 6(9), 831–839. <https://doi.org/10.1038/ncb1164>
- Jirimutu, Wang, Z., Ding, G., Chen, G., Sun, Y., Sun, Z., Zhang, H., Wang, L., Hasi, S., Zhang, Y., Li, J., Shi, Y., Xu, Z., He, C., Yu, S., Li, S., Zhang, W., Batmunkh, M., Batsukh, T. S., ... Meng, H. (2012). Genome sequences of wild and domestic bactrian camels. *Nature Communication*, 3, 1202. <https://doi.org/10.1038/ncomms2192>
- Kido, M., Du, L., Sullivan, C. C., Li, X., Deutsch, R., Jamieson, S. W., & Thistlethwaite, P. A. (2005). Hypoxia-inducible factor 1- $\alpha$  reduces infarction and attenuates progression of cardiac dysfunction after myocardial infarction in the mouse. *Journal of the American College Cardiology*, 46(11), 2116–2124. <https://doi.org/10.1016/j.jacc.2005.08.045>
- Lei, C. Z., Chen, H., Zhang, H. C., Cai, X., Liu, R. Y., Luo, L. Y., Wang, C. F., Zhang, W., Ge, Q. L., Zhang, R. F., Lan, X. Y., & Sun, W. B. (2006). Origin and phylogeographical structure of Chinese cattle. *Animal Genetics*, 37(6), 579–582. <https://doi.org/10.1111/j.1365-2052.2006.01524.x>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>
- Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marschall, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli, K.-P., ... Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, 479(7373), 359–364. <https://doi.org/10.1038/nature10574>
- Lorenzo, F. R., Huff, C., Myllymäki, M., Olenchock, B., Swierczek, S., Tashi, T., Gordeuk, V., Wuren, T., Ri-Li, G. E., McClain, D. A., Khan, T. M., Koul, P. A., Guchhait, P., Salama, M. E., Xing, J., Semenza, G. L., Liberzon, E., Wilson, A., Simonson, T. S., ... Prchal, J. T. (2014). A genetic mechanism for Tibetan high-altitude adaptation. *Nature Genetics*, 46(9), 951–956. <https://doi.org/10.1038/ng.3067>
- Maguire, J. J., Johnson, C. M., Mockridge, J. W., & Davenport, A. P. (1997). Endothelin converting enzyme (ECE) activity in human vascular smooth muscle. *British Journal of Pharmacology*, 122(8), 1647–1654. <https://doi.org/10.1038/sj.bjp.0701564>
- Mauer, J., Chaurasia, B., Goldau, J., Vogt, M. C., Ruud, J., Nguyen, K. D., Theurich, S., Hausen, A. C., Schmitz, J., Brönneke, H. S., Estevez, E., Allen, T. L., Mesaros, A., Partridge, L., Febbraio, M. A., Chawla, A., Wunderlich, F. T., & Brüning, J. C. (2014). Signaling by IL-6 promotes alternative activation of macrophages to limit endotoxemia and obesity-associated resistance to insulin. *Nature Immunology*, 15(5), 423–430. <https://doi.org/10.1038/ni.2865>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524>
- Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H., Zhao, C., Zhao, S., Song, J., Guang, X., Liu, G. E., Li, A., Wu, X., Wang, C., Fang, X., Zhao, X., Smith, S. B., Yang, W., Tian, W., ... Zhan, L. (2017). Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Molecular Biology and Evolution*, 35(3), 688–699. <https://doi.org/10.1093/molbev/msx322>
- Miao, B. P., Wang, Z., & Li, Y. X. (2017). Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the gray wolf from the Tibetan Plateau. *Molecular Biology and Evolution*, 34(3), 734–743. <https://doi.org/10.1093/molbev/msw274>
- Moncada, S., & Higgs, E. A. (1991). Endogenous nitric oxide: Physiology, pathology and clinical relevance. *European Journal of Clinical Investigation*, 21(4), 361–374. <https://doi.org/10.1111/j.1365-2362.1991.tb01383.x>
- Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

- Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, *541*(3), 302–310. <https://doi.org/10.1038/nature21347>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. <https://doi.org/10.1086/519795>
- Qiu, Q., Wang, L., Wang, K., Yang, Y., Ma, T., Wang, Z., Zhang, X., Ni, Z., Hou, F., Long, R., Abbott, R., Lenstra, J., & Liu, J. (2015). Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nature Communication*, *6*, 10283. <https://doi.org/10.1038/ncomms10283>
- Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D. M., Auvil, L., Capitanu, B., Ma, J., Lewin, H. A., Qian, X., Lang, Y., Zhou, R., Wang, L., Wang, K., ... Liu, J. (2012). The yak genome and adaptation to life at high altitude. *Nature Genetics*, *44*(8), 946–949. <https://doi.org/10.1038/ng.2343>
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489–494. <https://doi.org/10.1038/nature08365>
- Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925. <https://doi.org/10.1038/ng.3015>
- Song, C., Chang, X. J., Bean, K. M., Proia, M. S., Knopf, J. L., & Kriz, R. W. (1999). Molecular characterization of cytosolic phospholipase A2-beta. *Journal of Biological Chemistry*, *274*(24), 17063–17067. <https://doi.org/10.1074/jbc.274.24.17063>
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.
- Tekin, D., Dursun, A. D., & Xi, L. (2010). Hypoxia inducible factor 1 (HIF-1) and cardioprotection. *Acta Pharmacologica Sinica*, *31*(9), 1085–1094. <https://doi.org/10.1038/aps.2010.132>
- Thompson, L., Dong, Y., & Evans, L. (2009). Chronic hypoxia increases inducible NOS-derived nitric oxide in fetal guinea pig hearts. *Pediatric Research*, *65*(2), 188–192. <https://doi.org/10.1203/PDR.0b013e31818d6ad0>
- Verma, P., Sharma, A., Sodhi, M., Thakur, K., Kataria, R. S., Niranjani, S. K., Bharti, V. K., Kumar, P., Giri, A., Kalia, S., & Mukesh, M. (2018). Transcriptome analysis of circulating PBMCs to understand mechanism of high altitude adaptation in native cattle of Ladakh region. *Scientific Reports*, *8*(1), 7681. <https://doi.org/10.1038/s41598-018-25736-7>
- Wang, G. D., Xie, H. B., Peng, M. S., Irwin, D., & Zhang, Y. P. (2014). Domestication genomics: Evidence from animals. *Annual Review of Animal Biosciences*, *2*(1), 65–84. <https://doi.org/10.1146/annurev-animal-022513-114129>
- Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164. <https://doi.org/10.1093/nar/gkq603>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Wu, D.-D., Ding, X.-D., Wang, S., Wójcik, J. M., Zhang, Y. I., Tokarska, M., Li, Y., Wang, M.-S., Faruque, O., Nielsen, R., Zhang, Q., & Zhang, Y.-P. (2018). Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nature Ecology & Evolution*, *2*(7), 1139–1145. <https://doi.org/10.1038/s41559-018-0562-y>
- Xi, H., Chen, Z., Li, W., Wen, Y., Zhang, H., Xiao, Y., Liu, S., Pei, L., Zhang, M., Lv, P. O., Ren, F. U., Huang, K., Ye, L., Li, C., & Zhao, L. (2016). Chest circumference and sitting height among children and adolescents from Lhasa, Tibet compared to other high altitude populations. *American Journal of Human Biology*, *28*(2), 197–202. <https://doi.org/10.1002/ajhb.22772>
- Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., Yang, L., Pan, X., Wang, J., Shen, Y., Wu, B., Wang, H., & Jin, L. (2011). A genome-wide search for signals of high-altitude adaptation in Tibetans. *Molecular Biology and Evolution*, *28*(2), 1003–1011. <https://doi.org/10.1093/molbev/msq277>
- Yang, J. I., Li, W.-R., Lv, F.-H., He, S.-G., Tian, S.-L., Peng, W.-F., Sun, Y.-W., Zhao, Y.-X., Tu, X.-L., Zhang, M., Xie, X.-L., Wang, Y.-T., Li, J.-Q., Liu, Y.-G., Shen, Z.-Q., Wang, F., Liu, G.-J., Lu, H.-F., Kantanen, J., ... Liu, M.-J. (2016). Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Molecular Biology and Evolution*, *33*(10), 2576–2592. <https://doi.org/10.1093/molbev/msw129>
- Yang, Y., Fu, Q., Wang, X., Liu, Y., Zeng, Q., Li, Y., Gao, S., Bao, L., Liu, S., Gao, D., Dunham, R., & Liu, Z. (2018). Comparative transcriptome analysis of the swimbladder reveals expression signatures in response to low oxygen stress in channel catfish, *Ictalurus punctatus*. *Physiological Genomics*, *50*(8), 636–647. <https://doi.org/10.1152/physiolgenomics.00125.2017>
- Yuan, T., Wang, Y., Zhao, Z. J., & Gu, H. (2010). Protein-tyrosine phosphatase *PTPN9* negatively regulates ErbB2 and epidermal growth factor receptor signaling in breast cancer cells. *Journal of Biological Chemistry*, *285*(20), 14861–14870. <https://doi.org/10.1074/jbc.M109.099879>
- Zhang, H., Pajjmans, J. L. A., Chang, F., Wu, X., Chen, G., Lei, C., Yang, X., Wei, Z., Bradley, D. G., Orlando, L., O'Connor, T., & Hofreiter, M. (2013). Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nature Communication*, *4*, 2755. <https://doi.org/10.1038/ncomms3755>
- Zhang, Y. (2011). *Animal genetic resources in China*. Bovines [M]. China Agricultural Press.
- Zheng, B., Xu, Q., & Shen, Y. (2002). The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: Review and speculation. *Quaternary International*, *97–98*, 93–101. [https://doi.org/10.1016/s1040-6182\(02\)00054-x](https://doi.org/10.1016/s1040-6182(02)00054-x)
- Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassel, C. P., Sonstegard, T. S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J. A., & Salzberg, S. L. (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biology*, *10*(4), R42. <https://doi.org/10.1186/gb-2009-10-4-r42>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Liu X, Li Z, Yan Y, et al. Selection and introgression facilitated the adaptation of Chinese native endangered cattle in extreme environments. *Evol Appl*. 2021;14:860–873. <https://doi.org/10.1111/eva.13168>