BMC Genomics

RESEARCH

Open Access

Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity characteristics in three Chinese cattle breeds

Xiaoyun Wu[1], Jie Pei[1], Lin Xiong[1], Qianyun Ge[1], Pengjia Bao[1], Chunnian Liang[1], Ping Yan[1,2,3*] and Xian Guo[1*]

Abstract

Background Cattle have evolved genetic adaptations to a diverse range of agroecological zones, such as plateaus and arid zones. However, little is known about its genetic basis of adaptation to harsh environments within a short period of time after domestication. Here, we analyzed whole-genome sequence data from three indigenous cattle breeds (Anxi, Qaidam and Zhangmu) in northwest China and five worldwide cattle breeds (Angus, Holstein, Jersey, Gir and N'Dama) to explore their genetic composition and identify selective sweeps in the Chinese cattle breeds.

Results Analyses of phylogenetic and population structure revealed that three indigenous cattle breeds share genomic components from *Bos taurus* and *Bos indicus*. A novel set of candidate genes was identified through comparative genomic analyses of cattle from contrasting environments based on SNP and copy number variation (CNV) data. These candidate genes are potentially associated with adaptive phenotypes, including high-altitude adaptability (e.g., ANGPT1, PPARGC1A, RORA), cold climate adaptation (e.g., TSHR, PRKG, OXCT1), and dryland adaptation (e.g., PLEKHA7, NFATC1, PLCB1).

Conclusions This study unravels the unique adaptive diversity of three Chinese indigenous cattle breeds, providing a valuable resource for future research on sustainable livestock breeding strategies to response to climate change.

Keywords Chinese cattle, Extreme environment, Adaptation, Genome, Selection signature

Introduction

Cattle, as an invaluable livestock species worldwide, are raised for meat, milk, leather, transportation and cultivation. Due to their ability to convert poor quality forage into high-quality meat and milk products, cattle provide an important source of nutrition and an economic livelihood for nearly 6.6 billion of people worldwide [1]. There are two main lineages of modern-day cattle, the humpless taurine breeds (*Bos taurus*) and the humped zebu (*Bos indicus*), which originated from separate domestication events in different locations in the Fertile Crescent around 10,000 years ago and on the Indian subcontinent

*Correspondence:
Ping Yan
pingyanlz@163.com
Xian Guo
guoxian@caas.cn
[1]Key Laboratory of Animal Genetics and Breeding on Tibetan Plateau, Ministry of Agriculture and Rural Affairs, Key Laboratory of Yak Breeding Engineering, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou 730050, P.R. China
[2]Institute of Western Agriculture, The Chinese Academy of Agricultural Sciences, Changji 831100, China
[3]Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China

approximately 8000 years ago [2, 3]. Following domestication, hundreds of *taurine* and *indicus* cattle breeds have been created by subtle combinations of natural and human-mediated artificial selection, exhibiting a range of environmental adaptations and agricultural traits, such as milk production and meat performance. Hybridization and genetic introgression with other bovids have also contributed to increased genetic diversity within Bovinae [4].

China has a rich resource of indigenous cattle breeds, and these breeds can be classified into three groups based on their ecological features, and exhibit phenotypic variability and diversified environmental adaptability [5]. There are some unique indigenous cattle breeds distributed in northwest China, which have genetically evolved to adapt to extreme environments such as deserts and plateaus. The unique genetic composition and adaptations of these cattle breeds represent a critical genetic resource in the face of climate change and increased food requirements in the future. During recent decades, these indigenous cattle breeds have been in continuous decline due to changing production systems, as well as indiscriminate crossbreeding with other breeds. Thus, a deeper understanding of the genetic basis of the indigenous cattle breeds in northwest China is essential for developing climate-resilient breeds response to future climate changes and preserve valuable cattle genetic resources.

Exploration genomic divergence can elucidate the genetic basis of adaptation to various environments and identify genetic variants with significant functions. Whole genome sequencing technologies have been widely used to characterize adaptive genetic variation in various species inhabiting harsh or extreme environments. Candidate genes contributing to environmental adaptation have been identified using single nucleotide polymorphism (SNP) markers in humans [6], as well as in various farm animals such as yaks [7], Tibetan chickens [8], cattle [9], pigs [10], and sheep [11]. Copy number variation (CNV) is a class of genetic variation that has remarkable effects on gene expression through disrupting coding sequences, altering genedosage or position effects, and these effects could contribute to phenotypic diversity [12]. In recent years, emerging evidence has suggested that CNVs contribute to the adaptation of cattle to the harsh environments [13].

Northwestern China is characterized by a diverse climate range that significantly shapes its ecological and biological systems. Some local cattle breeds in northwestern China evolved to adapt to various extreme environments during domestication. For example, the Qaidam (QDM) and Zhangmu (ZM) cattle inhabit the Tibetan Plateau at altitudes of 3000–5000 m, where they are well adapted to the cold, hypoxic environment, and they are likely under selection for high-altitude adaptation. Meanwhile, Anxi cattle (AX) lives in the Gobi Desert, where they exhibit tolerance to cold, arid environments and poor-quality forage. In this study, we sequenced the whole genome of seven AX cattle and downloaded the genome sequence data of 57 cattle from seven other breeds (QDM, ZM, Angus, Holstein, Jersey, Gir and N'Dama) to explore the molecular mechanisms underlying the genome-wide adaptive response of three Chinese local cattle breeds to challenges of extreme environments.

Results

Sequencing and variation calling

We identified 33,588,078 SNPs by analyzing the whole-genome sequence data of 21 cattle from three Chinese breeds and publicly available genomes of 43 individuals from seven other cattle breeds sourced from NCBI. A total of 7,400,753, 7,237,974 and 8,770,574 SNPs were identified in the AX, QDM, and ZM cattle, respectively. ZM cattle exhibited the highest number of SNPs among all breeds. The Ts/Tv ratio, representing the average transition to transversion ratio, exhibits a similar pattern across different breeds. In European *taurine* cattle, the ratio is 2.32, whereas in Chinese indigenous cattle breeds, it is 2.34 (Additional file 1, Table S1). According to ANNOVAR annotation, the majority of identified SNPs were located in intergenic regions (21,423,034, 59.04%), followed by those in the intronic (13,648,767, 37.61%), exonic (426,433, 1.17%) and 3'UTR (270,921, 0.75%) regions (Additional file 1, Table S2).

In addition, a comprehensive analysis of CNV in the cattle genome was conducted in this study. A total of 14,117 copy number variation regions (CNVRs) were detected in 64 cattle, covering a total length of 62.01 Mb, which represents 2.36% of the ARS-UCD1.2 genome (Additional file 1, Table S3). CNVRs were categorized into three types, including deletion, duplication and both. The lengths of CNVRs range from 1.60 kb to 475.60 kb. The distribution of CNVRs across the chromosomes was not uniform. Chromosome 1 contained the highest number of CNVRs (846), while chromosome 28 contained the lowest number (215) (Additional file 1, Table S3).

Genomic diversity

Principal component analysis (PCA) analysis was conducted on subsets of genotype data for autosomal SNPs. In spite of the lack of breed membership, the analysis showed a clear structure since animals of the same breed were clustered together. The first two principal coordinates from multidimensional scaling (MDS) explained 17.71% and 8.90% of the total genomic variance, respectively (Fig. 1a). The first principal component (PC1) was driven by differences between *taurin*e and *indicine* cattle. The PC2 separates N'Dama and those of China and

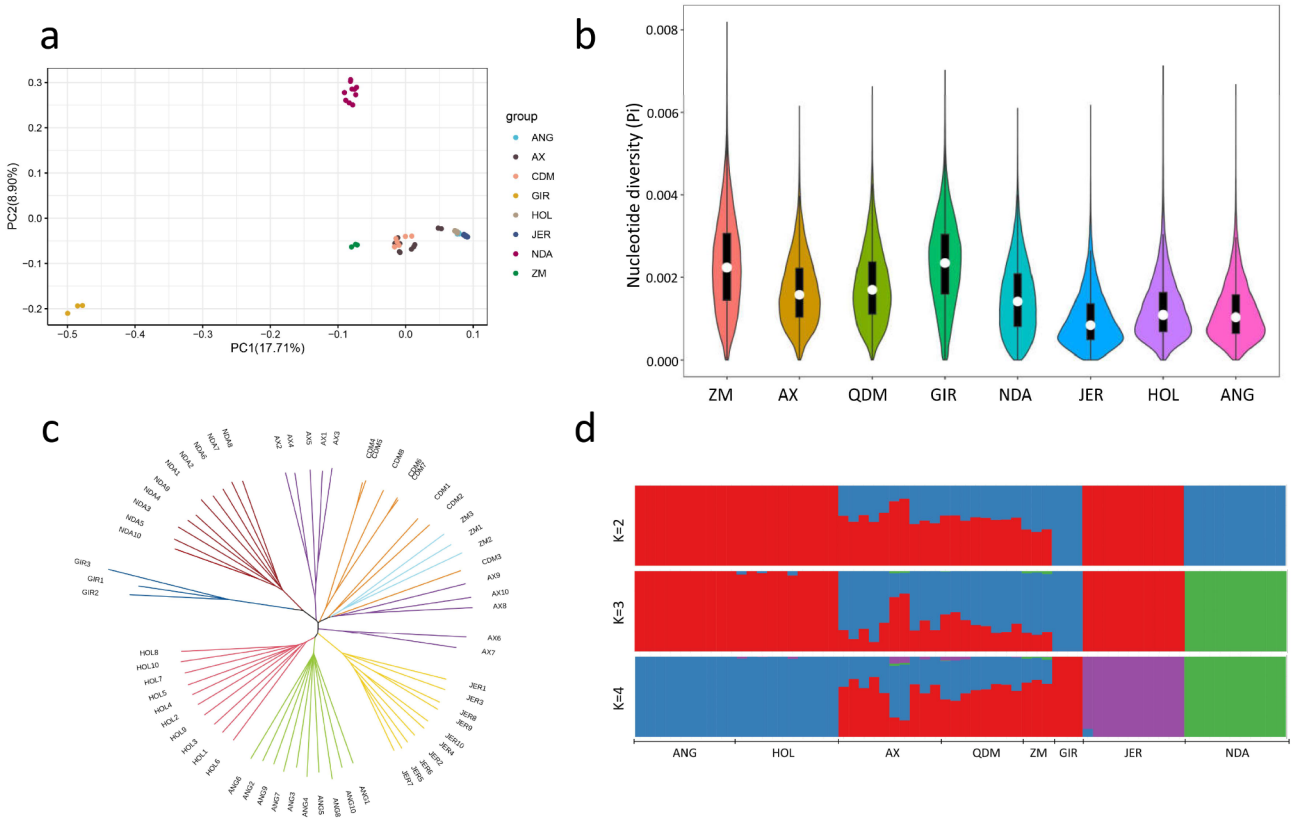

Fig. 1 Population structure and relationships among eight cattle breed. (**a**) PCA of genotypes for three native Chinese breeds and additional cattle breeds. (**b**) Genome-wide nucleotide diversity distribution, (**c**) Neighbor-joining tree of the relationships between the eight cattle breeds. (**d**) Population structure analysis of each sample assuming different number of ancestral breeds (K = 2, 3, and 4)

Europe. On a genome-wide window scale of 10 Mb, ZM and GIR cattle exhibited the highest nucleotide diversity, whereas the JER cattle displayed the lowest (Fig. 1b). All Chinese indigenous breeds exhibited higher levels of nucleotide diversity than the commercial European breeds. The neighbor-joining (NJ) tree showed a clear division between *taurine* and *indicine* cattle. The three Europe cattle breeds (ANG, JER and HOL) clustered together, while AX, QDM and ZM cattle clustered together, which are located at an intermediate position between *Bos taurus* and *Bos indicus* (Fig. 1c). Additionally, an admixture analysis was performed to infer the proportions of ancestry in the subsample used in this study with various numbers of clusters (K), corresponding to hypothetical ancestral populations. In this study, we found that K = 4 was the most probable number of genetically distinct clusters. When K = 4, AX, QDM and ZM cattle displayed significant evidence of genetic heterogeneity, sharing a common genetic heritage with both *taurine* and *indicine* cattle (Fig. 1d, Additional file 2, Table S1).

Selective signatures associated with high-altitude adaptation

To investigate the adaptive mechanisms of cattle in plateau environment, we compared the high-altitude group with low-altitude group to identify selective signals. The 5% of windows with the highest Fst and π-ratio values were regarded as potential selected windows. For Tajima's D, the top 5% of windows with significantly negative Tajima's D values were considered as candidate sweeps. The regions identified by at least two methods were ascertained as the final selective sweeps to minimize method-specific bias. Totally, we identified 2,218 and 1,875 candidate genes from Fst (Fst > 0.166) and π-ratio (π-ratio > 1.210), respectively (Fig. 2a and b; Additional file 1, Table S4 and S5). In addition, Tajima D (Tajima's D value < -1.629) analysis identified 1,958 candidate genes in high-altitude group (Fig. 2c; Additional file 1, Table S6). Finally, 624 genes (Fig. 2d) were identified as positively selected genes (PSG) in high-altitude group. We found that PSGs were significantly enriched in multiple GO terms and pathways associated with highaltitude adaptation, such as regulation of heart rate by cardiac conduction (GO:0086091), heart contraction (GO:0060047), blood vessel morphogenesis (GO:0008015), sprouting

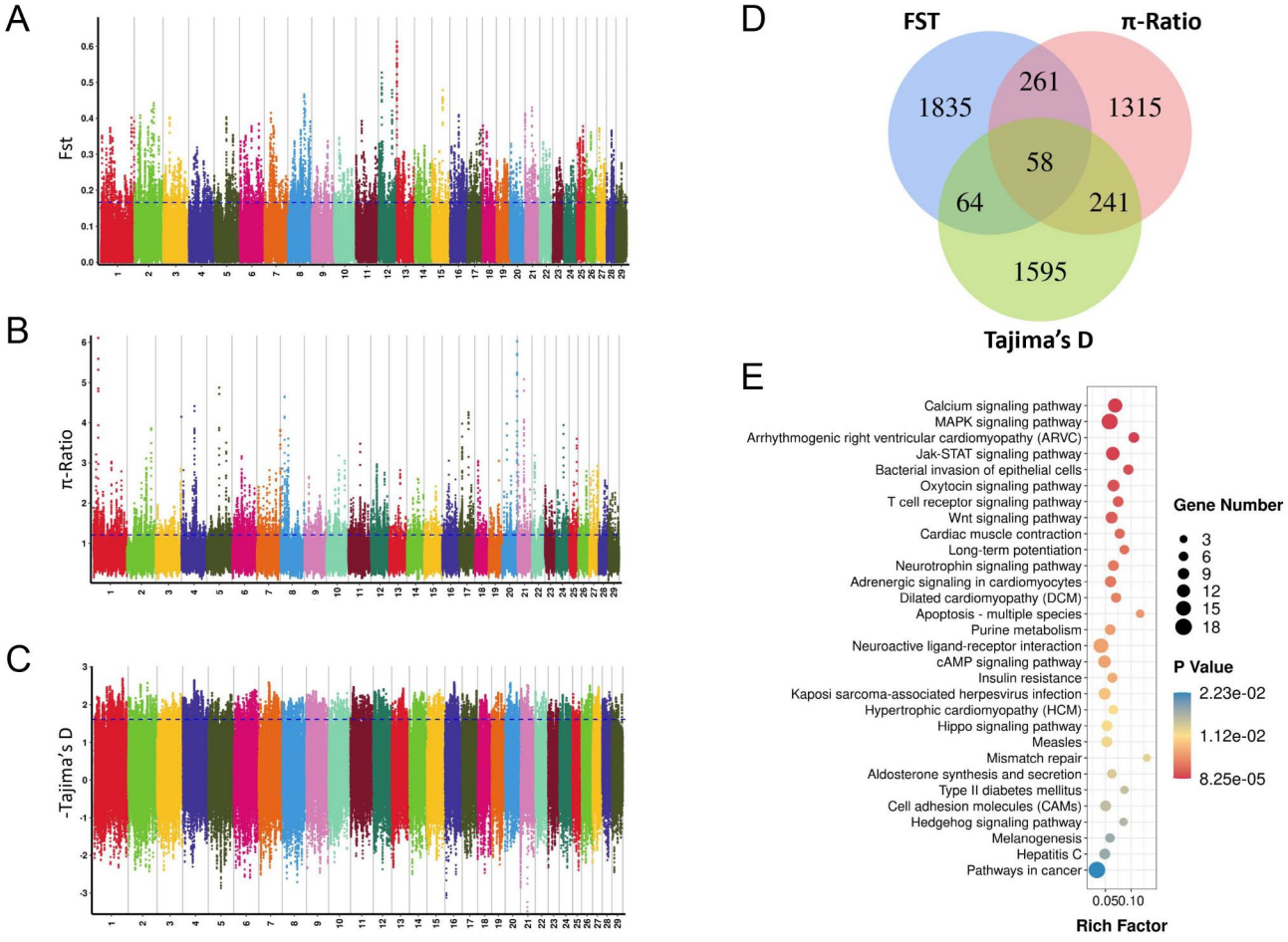

Fig. 2 Identification of candidate genes for high-altitude adaptation. (**a**-**c**) Manhattan plot of the genome-wide distribution of Fst, *π*-ratio and Tajima'D using 100 kb windows size and 10 kb step size, respectively. (**d**) Number of candidate genes identified in highland group by the three methods listed in each of the Venn diagram components. (**e**) KEGG enrichment analysis for the identified candidate genes

angiogenesis (GO:0002040), response to ultraviolet (UV) (GO:0009411) (Additional file 1, Table S7), calcium signaling pathway (bta04020), cardiac muscle contraction (bta04260) and purine metabolism (bta00230) (Fig. 2e; Additional file 1, Table S8). These GO terms and pathways are associated with adaptation to plateau environments of cattle because they indicate potential improvements in either the ability to oxygen transportation or defense against UV.

Based on the Vst analysis of the CNVRs between the two groups, 707 CNVRs with Vst > 0.320 (top 5%) were screened, overlapping with 215 coding genes (Additional file 1, Table S9). Of these PSGs, 26 were overlapped with SNP-based PSGs (Additional file 1, Table S10). Some PSGs were significant enrichment in hair cycle (GO:0042633), regulation of heart rate (GO:0002027), vasculature development (GO:0001944), heart development (GO:0007507), and lung development (GO:0030324) (Additional file 1, Table S11).

Selective signatures associated with cold adaptation

To understand the adaptive mechanisms in frigid environments, we compared the cold-adapted group (AX, QDM, and ZM cattle) to NDA, which is known for its adaptation to hot climatic conditions in Africa. We identified 2,166, 1,985 and 1,979 candidate genes using Fst, π-ratio and Tajima'D methods, respectively (Additional file 1, Table S12-S14). In total, 1,029 candidate genes were identified as PSGs in the cold-adapted group. Analysis of the PSGs revealed several categories that were relevant to adaptation to cold environments. These categories, including response to cold (GO:0009409), forebrain development (GO:0030900), lipid transport (GO:0006869), blood vessel development (GO:0001568), regulation of carbohydrate metabolic process (GO:0006109), kidney development (GO:0001822), and glycerolipid metabolic process (GO:0045017) appeared to be functionally relevant to cold adaptation (Additional file 1, Table S15). KEGG pathway analysis of the PSGs revealed significant enrichment in 57 pathways, such as metabolic pathways (bta01100), nitrogen metabolism

(bta00910), renin secretion (bta00910), regulation of lipolysis in adipocytesn signaling pathway (bta04923), and glycerolipid metabolism (bta00561) (Additional file 1, Table S16).

We also identified 706 potentially selected CNVRs in cold-adapted group, containing 174 annotated protein coding genes (Fig. 3a; Additional file 1, Table S17). Among these PSGs, 35 were overlapped with SNP-based PSGs (Additional file 1, Table S18). KEGG pathway analysis revealed that 98 pathways were significantly enriched (Fig. 3b; Additional file 1, Table S19). Among these pathways, Wnt signaling pathway (bta04310), calcium signaling pathway (bta04020), valine, leucine and isoleucine degradation (bta00280), metabolic pathways (bta01100), VEGF signaling pathway (bta04370), and oxytocin signaling pathway (bta04921) appeared to be functionally relevant to the cold adaptation.

Selective signatures associated with adaptation to dryland stress

To detect potential selective sweeps associated with arid environments, we compared the dryland-adapted group with control group (including ZM and GIR cattle). Regions with the top 5% of Fst values (Fst > 0.31) and the π-ratio values (π-ratio > 3.33) were selected as candidate regions potentially associated with adaptation in the dryland-adapted group (Fig. 4a; Additional file 1, Table S20 and S21). A total of 854 genes with strong selective sweep signals were identified in the dryland-adapted group. In addition, the Tajima'D (Tajima's D value < -1.629) analysis identified 2,105 candidate genes in the dryland-adapted group (Additional file 1, Table S22). Totally, 1,548 candidate genes were identified as PSGs in the dryland-adapted group (Fig. 4b).

In particular, pleckstrin homology domain containing A7 (PLEKHA7), a gene involved in the regulation of intracellular calcium, showed strong positive selection in the dryland group (Fig. 4c). GO and KEGG pathway analyses of the PSGs identified 1,557 and 74 significantly enriched GO terms and pathways, respectively (Additional file 1, Table S23 and S24).

In the selective sweep analysis of CNVRs, we identified 204 candidate selected CNVRs in dryland-adapted group, which overlapped with 173 coding genes (Additional file 1, Table S25). Of these PSGs, 29 were overlapped with SNP-based PSGs (Additional file 1, Table S26). The KEGG analysis revealed that some PSGs were significantly enriched in pathways associated with drought tolerance, such as salivary secretion, sphingolipid signaling pathway, and metabolic pathways (Additional file 1, Table S27).

Discussion

The unique genetic composition and adaptation of cattle breeds in northwest China reflect a significant cultural legacy and genetic variation associated with traits related to environmental adaptation. Further genetic studies on these cattle breeds would provide insights into improving them in response to future climate changes. In this study, we investigated the genetic composition and identified valuable SNP and CNVs associated with local adaptation and in three indigenous Chinese cattle breeds using genome sequence data from 64 individuals.

Nucleotide diversity was used to estimate the level of polymorphism within a population [14]. In this study, the commercial European breeds (HOL, ANG, and JER) exhibited lower levels of nucleotide diversity compared to other cattle breeds, consistent with previous research [15]. The reduced level of nucleotide diversity in commercial European breeds may be a consequence of high-intensity human-mediated selection for production traits over a few generations or a low effective population size dominated by genetic drift. In contrast, the ZM cattle exhibit the highest nucleotide diversity. The ZM cattle is a hybrid breed, originally comprising *taurine* cattle crossed with *indicine* cattle as well as yak [16]. The relatively high

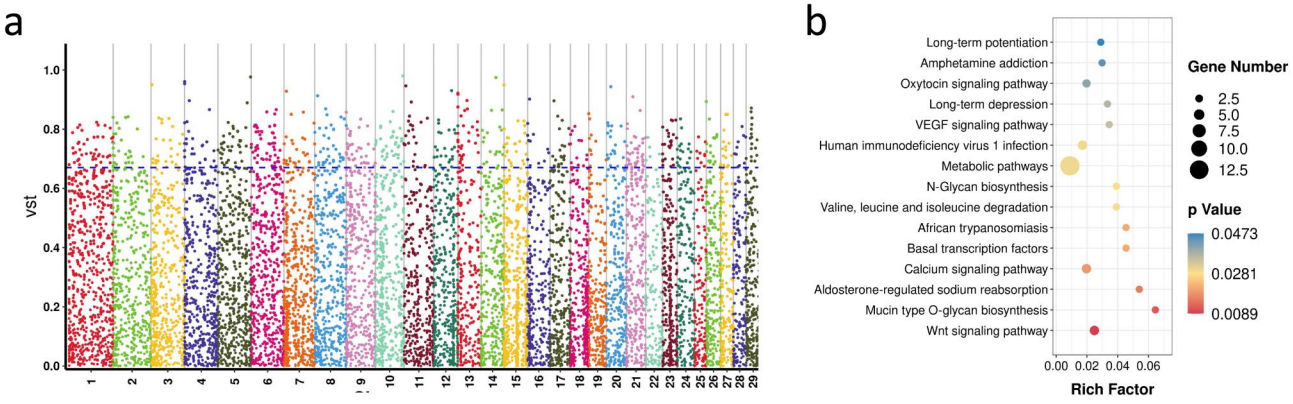

Fig. 3 Identification of candidate genes for cold enviroment adaptation based on CNV. (**a**) The Vst values of all the CNVRs in cold-adapted group and control group. (**b**) KEGG enrichment analysis for the identified candidate genes

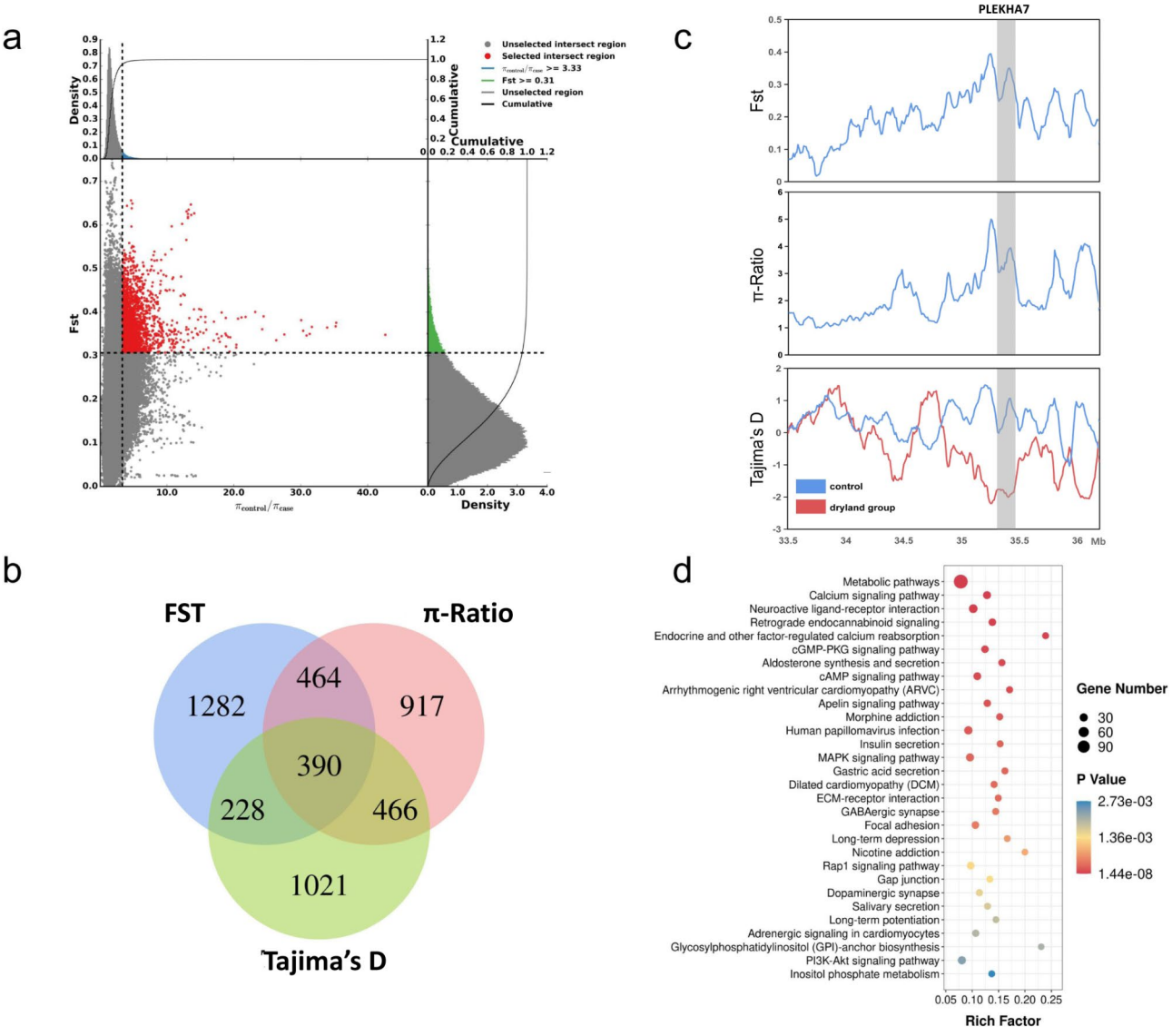

Fig. 4 Genomic regions with strong selective signals in dryland group. (**a**) Distribution of log2 (π-ratio [θπ,Others/ θπ, dryland group]) (top 5% outliers, π-ratio > 3.33) and Fst values (top 5% outliers, Fst > 0.31), which are calculated in 100 kb windows sliding in 10 kb steps. (**b**) Venn diagram of candidate genes screened by Fst, π-ratio and Tajima'D in dryland group. (**c**) Selective signals around the PLEKHA7 gene. (**d**) KEGG enrichment analysis for the identified candidate genes

genetic diversity observed in ZM cattle might result from introgression of *indicine* cattle and yak. Ancestry component analysis can indicate the degree of genetic information exchange. Additionally, three indigenous Chinese cattle breeds share genome ancestry with *taurine* and *indicine* lineage, which is consistent with previous studies [3, 5].

Adaptive mechanisms in plateau environments

The Tibetan Plateau, known as the "roof of the world", hosts extreme environments conditions, including low oxygen levels, low temperatures, and high UV radiation levels. Over their evolutionary histories, numerous mammal species inhabiting this plateau have developed diverse adaptive traits that enhance their capacity to survive in the high-altitude environment, such as large lungs and hearts, high blood flow, and increased energy metabolism. In recent years, there has been increasing interest in identifying the genetic factors that play a crucial role in high-altitude adaptation in Tibetan people [17, 18], yak [7], Tibetan antelope [19], as well as Tibetan pigs [10].

In this study, we uncovered a series of genes involved in a several GO terms and signaling pathways related to cardiovascular system and energy metabolism, such as blood vessel morphogenesis (e.g., *ANGPT1*, *MMRN2* and *BMPER*), sprouting angiogenesis (e.g., *PPARGC1A*, *BMPR1B* and *IL6ST*), regulation of heart rate by cardiac conduction (e.g., *RECK*, *LRP6* and *CTNNBIP1*,

and purine metabolism (e.g., *FHIT*, *PGM2* and *ADK*). Hypoxia inducible factors (HIFs) consist of a HIF-1α subunit that is regulated by oxygen and a HIF-1β subunit that is expressed continuously. HIFs are essential for regulating high-altitude adaptation in humans and animals. Among genes with evidence of positive selection, angiopoietin 1 (ANGPT1) belongs to the angiopoietin family and is a major downstream target of HIF-1. *ANGPT1* has been shown to be involved in angiogenesis and increased tissue vascularization, thereby enhancing oxygen delivery [20]. Recent studies have shown that *ANGPT1* has evolved under positive selection for hypoxia adaptation in native Tibetan humans [21] and Himalayan wolves [22]. Peroxisome proliferator-activated receptor gamma coactivator 1-alpha (*PPARGC1A*) is a crucial transcriptional coactivator involved in controlling mitochondrial biogenesis and determining muscle fiber type [23]. An increased number of slow muscle fibers has been reported to enhance metabolic capacity and oxygen delivery in plateau animals [24, 25]. The *PPARGC1A* gene has been positively selected in high-altitude group, indicating that enhancing oxygen-diffusing capacity through an increase in capillarity and oxidative fiber number may be a crucial way for cattle to respond to low oxygen pressure. In addition to hypoxia, long-term exposure to strong solar and ultraviolet radiation (UV) radiation threatens the survival of high-altitude species, as UV radiation could induce direct damage to DNA. Response to UV and DNA repair is one of the significantly enriched GO terms in our analyses of the high-altitude group, as well as from the studies of other species at high altitudes [26]. Combined with evidence from Tibetan chickens [27], American pikas [28], the highest-elevation frog [29], and snub-nosed monkeys [30], our results suggest that defense against high UV radiation is a potential adaptive mechanism for highland animals.

In addition to SNPs, we identified 215 coding genes that overlap with highly differentiated CNVRs between high-altitude group (QDM and ZM) and low-altitude group. Among them, several genes (e.g., *IGFBP5*, *LIF*, *PROX1*, *RORA* and *CDH13*) involved in vasculature development (e.g., *PROX1*, *RORA* and *CDH13*) and lung development (e.g., *IGFBP5*, *LIF*, and *SHH*) have well-established biological functions related to high-altitude adaptation. CNV in genes might cause changes in gene expression through a dosage effect. Insulin-like growth factor binding protein 5 (IGFBP5) plays a crucial role in key cellular adaption to angiogenesis and regulation of smooth muscle cell proliferation [31]. Over-expression of *IGFBP5* can lead to increases the expressions of angiogenic markers [32]. In our study, the *IGFBP5* gene exhibited a reduced copy number in low-altitude group compared to the high-altitude group, implying that the reduced dosage of the *IGFBP5* gene may reduce capacity to adapt to hypoxia through affecting the cardiovascular system of cattle. Leukemia inhibitory factor (*LIF*) is a pleiotropic cytokine belonging to the interleukin-6 cytokine family, whose expression is highest in mouse lung at birth. A previous study has shown that LIF and IGF-I cooperatively regulate lung alveolar epithelium and vascular maturation [33]. The presence of multiple copies at the LIF gene locus was observed only in high-altitude group in the current study, implying that the increased dosage of *LIF* may benefit the adaptation to high altitude by mediating the lung development.

Adaptive mechanisms in cold environments

Ambient temperature is one of the important non-biological challenges for livestock that affects their feed efficiency, health and reproduction. Homeothermy responds to cold stress through various adaptive mechanisms, such as increasing production of heat by enhancing the basal metabolic rate and insulation [34]. Cold is one of the most significant environmental stress factors for the cattle breeds from northwest China. In this study, 219 PSGs corresponding to selective sweeps were identified in the cold-adapted group. GO analysis identified a significant enrichment of PSGs involved in biological processes that contribute to the maintenance of constant body temperature during cold stress. These processes were related to glycerolipid metabolic process, lipid metabolic process, kidney development, blood vessel development, and forebrain development, among others. This finding is consistent with previous studies suggesting that cold tolerance is a multifaceted process involving multiple biological processes, such as (1) energy metabolism, (2) nervous system development, (3) blood circulation [34, 35].

Of all PSGs, we identified 17 PSGs associated with lipid metabolism and thermogenesis, including thyroid stimulating hormone receptor (*TSHR*), cGMP-dependent protein kinase (*PRKG1*), lipoprotein lipase (*LPL*), monoacylglycerol O-acyltransferase 1 (*MOGAT1*), acyl-CoA synthetase long-chain family member 3 (*ACSL3*), and glycerol kinase (*GK*). This finding is consistent with previous research highlighting the significance of fat metabolism in cold adaptation in chickens [36], Siberians [37], Yakutian horses [38], and woolly mammoths [39]. It is also worth noting that *TSHR* has been reported to be related to regulation of metabolism and photoperiod control of reproductive cycles in vertebrates [40]. TSHR belongs to the G protein-coupled receptor (GPCR) superfamily and plays an important role in thyroid development and function [41]. Recent studies have shown that *TSHR* plays a role in regulating energy balance, metabolism, and thermoregulation [42, 43]. The *PRKG1* gene plays a crucial role in regulating cardiovascular homeostasis [44]. It has been identified as a candidate gene for feed conversion efficiency, intramuscular

fat content, and meat tenderness in cattle [45]. It was also found that the *PRKG1* gene was subjected to selection in several species such as sheep [11], the Amur Tiger [46], and Siberian [37], which inhabit cold environments. Therefore, we propose that these genes associated with energy metabolism may play a role in the adaptation of cattle to low-temperature environments.

A total of 170 genes were identified as potential PSGs according to Vst test on the copy number of CNVRs. Some of these PSGs were enriched in GO terms related to cold tolerance, such as blood circulation, kidney development, and glycerolipid metabolic process, suggesting that CNVs contribute to cold adaptation in cattle. For instance the 3-Oxoacid CoA-Transferase 1 (*OXCT1*) gene is involved in extrahepatic ketone body catabolism and is abundantly expressed in brown adipocytes. It has been reported that *OXCT1* is involved in the maintenance of body core temperature in mice [47]. Phospholipase C epsilon 1 (PLCE1), a member of the phospholipase family, plays a crucial role in intracellular signaling by catalyzing the hydrolysis of membrane phospholipids [48]. *PLCE1* has been identified as a candidate gene for climate adaptation in sheep [49, 50]. Both studies indicated that variations in *PLCE1* gene were associtaed with the mean diurnal temperature range. The presence of multiple copies of *OXCT1* and *PLCE1* genes were identified in cold-adapted group, suggesting that the increased dosage of *OXCT1* and *PLCE1* may benefit the adaptation to low-temperature environments.

Adaptive mechanisms in dryland

The Gobi region is characterized by arid conditions, low precipitation, high evaporation, intense solar radiation, and large temperature fluctuations. To cope with the challenges of arid environments, animals inhabiting Gobi region have evolved unique physiological traits, including enhanced use of metabolic water, reduced water losses via urine and feces, the ability to tolerate a high-salt diet without developing high blood pressure, and adaptive tolerance to starvation and dehydration [51]. In this study, PSGs in the dryland-adapted group were significantly enriched in metabolic pathways, pancreatic secretion, insulin secretion, and oxytocin signaling pathway. These pathways are crucial for adaptability to water deprivation and poor-quality forages in dryland region. Similar findings have been reported for the Bactrian camel [52] and sheep [11], suggesting potential convergent evolution in mammals adapting to arid desert environments. In addition, we found that PSGs in the dryland-adapted group were significantly enriched in some well-known GO terms and pathways related to adaptation to a dryland environment, such as renal system development, regulation of sodium ion transport, and intracellular calcium ion homeostasis, visual system development, water homeostasis, regulation of water loss via skin, vasopressin-regulated water reabsorption, and calcium signaling pathway. The kidney plays a vital role in regulating of water retention and reabsorption, and water retention in the kidney are associated with adaptation to starvation and dehydration in desert animals [51]. Hypertonicity regulates water balance and reabsorption in the kidney. Among the candidate genes for dryland stress adaptations, pleckstrin homology domain containing A7 (*PLEKHA7*) is a candidate gene for human hypertension, which plays a key role in reducing salt-sensitive hypertension and renal diseases in rats [53]. A recent study demonstrated that *PLEKHA7* is under positive selection for adaptation to arid environment in African sheep [54]. Another candidate gene worth noting, related to water reabsorption, is nuclear factor of activated T cells 1 (NFATC1). A recent study showed that *NFATC1* regulates the expression of genes associated with water reabsorption in the kidneys of the Yarkand hare [55]. Solar radiation and airborne dust are critical challenges in the desert environments, which can lead to a number of ophthalmic conditions and respiratory diseases [52]. We found that several PSGs were involved in visual system development and respiratory gaseous exchange by respiratory system, which may offer protection against solar and UV radiation. In addition, a few PSGs (e.g., *CMPK1*, *EIF2AK4*, *PARP1*, *BARD1* and *CHEK1*) in dryland-adapted group were enriched in responses to UV and DNA damage response. Intriguingly, Cytidine/uridine monophosphate kinase 1 (*CMPK1*) was previously reported to be under positive selection for adaptation to hot dryland environments in Egyptian sheep [56]. Eukaryotic translation initiation factor 2 alpha kinase 4 (*EIF2AK4*) was previously identified as a candidate gene for thermal tolerance traits in Chinese cattle [57]. These results provide evidence that AX and QDM cattle have undergone to adapt to the harsh environment of the Gobi desert.

In addition to SNPs, we identified 185 coding genes that overlapped with highly differentiated CNVRs between the dryland-adapted group with the control group. Interestingly, a CNVR on chromosome 10, which has the sixth highest Vst score and covers phospholipase C Beta 1 (*PLCB1*), a gene that was previously identified as being positively selected in sheep and goats adapting to dry-arid environments [58]. *PLCB1* was also identified as a PSG based on the SNP set. It hydrolyzes 1-phosphatidylinositol 4,5-bisphosphate into diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3). In this study, a reduced copy number of the *PLCB1* gene was observed only in the control group, suggesting that the *PLCB1* gene may be involved in adaptation to arid environment. ATP Binding Cassette Subfamily C Member 4(ABCC4) is a member of the ATP-binding cassette (ABC) superfamily

transporters, which is involved in cellular defense against oxidative stress [59]. A previous study demonstrated that the *ABCC4* gene is a candidate gene involved in heat stress and immune response in buffaloes. These results suggest a genetic basis for the ability of AX and QDM cattle to endure the challenges of the dryland. Moreover, it is recommended to validate the findings using alternative methods such as GWAS, candidate gene approach and gene expression analysis.

Conclusions

In this study, we identified numerous novel genes, crucial pathways and GO categories associated with local adaptations of cattle to plateau and arid environments. Notably, the candidate genes, GO categories and pathways were functionally implicated in cardiovascular system in plateau environment, energy metabolism in cold environments, and water reabsorption in the arid environment. These results expanded our knowledge of the genetic mechanisms underlying cattle adaptations to harsh environments.

Materials and methods

Sample collection and whole-genome sequencing

We sampled blood from a total of seven unrelated AX cattle in Guazhou County, Gansu Province. To ensure a representative sample, we consulted with livestock experts and owners of the cattle. After sample collection, the wound was disinfected with iodine volt to prevent infection. All operations of sample collection were allowed by the owner. No cattle were sacrificed for this study. The standard phenol-chloroform extraction protocol was employed for the extraction of blood genomic DNA. The quality and quantity of DNA were assessed using 0.8% agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer (NanoDrop, Wilmington, DE, USA). The Illumina TruSeq Nano DNA Library Prep Kit (Thermo Fisher Scientific, Waltham, MA, USA) was used to create paired-end sequencing libraries with an insert size of 150 bp. These libraries were subsequently sequenced on the Illumina HiSeq X Ten platform (Thermo Fisher Scientific). Moreover, resequencing data for eight cattle breeds, including AX ($n=3$), QDM ($n=8$), ZM ($n=3$), Holstein (HOL, $n=10$), Jersey (JER, $n=10$), Angus (ANG, $n=10$), Gir (GIR, $n=3$), and N'Dama (NDA, $n=10$) cattle were obtained from the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra) (Additional file 1 Table S28).

Sequence alignment and variation identification

Trimmomatic v0.36 was utilized to eliminate low-quality and adaptor contaminant sequences. The Burrows-Wheeler Alignment tool (BWA) version 0.7.5a (bwa-mem) [60] was used to align clean reads to the *Bos taurus* reference genome (ARS_UCD1.2). SAMtools [61] v1.1 was utilized for the conversion of SAM files into BAM files. Picard tools v1.1193 was used to identify and mark duplicate reads resulting from PCR amplification. The haplotypeCaller of the Genome Analysis Toolkit (GATK, version v3.6) [62] was used to call variants following the GATK best practices. The raw SNPs were filtered sing the "VariantFiltration" mode of GATK v4.1.2.0 based on the following criteria (-Window 4, -filter "QD < 4.0|| FS > 60.0|| MQ < 40.0", -G_filter "GQ < 20"). Finally, VCFtools [63] was employed to eliminate variants with missing genotypes exceeding 10% and minor allele frequencies (MAF) less than 0.05. Only biallelic SNPs were used for subsequent analysis. Copy number variations (CNVs) across autosomes were detected using the CNVcaller software [64]. The specific steps were in accordance with a reported study [65]. The raw CNVRs were merged into final CNVRs using the parameters (-f 0.1, -h 1, -r 0.5). Variation annotation and effect prediction were conducted using ANNOVAR [66] based on cattle reference genome and its annotation. To avoid gender bias, only variants on autosomal chromosomes were included in all analyses.

Population genetic analysis

Nucleotide diversity (in windows of 10 Mb) was initially estimated using VCFtools [63]. Plink 1.9 [67] was utilized to conduct the principal components analysis (PCA). ADMIXTURE (v1.3.0) [68] was used to estimate population structure and individual ancestry proportions with K ranging from 2 to 8 and 1000 bootstrap replicates. A phylogenetic tree was constructed using the neighbor-joining method was obtained using FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) based on the matrix of pairwise genetic distances among all samples.

Genome scanning for selection sweeps

To uncover genetic variants under selection, analyses were performed with the following four comparisons: (1) high-altitude group (QDM and ZM) versus low-altitude group (AX, HOL, ANG and JER), (2) cold-adapted group (QDM, AX and ZM) versus control group (NDA), (3) dryland-adapted group (QDM and AX) versus control group (ZM and GIR) (Additional file 1 Table S29).

For SNPs, we employed three distinct statistics (Fst, Tajima's D, and π-ratio) to identify positive selection signals. We utilized the VCFtools [63] to calculate Fst, π-ratio, and Tajima's D value using a sliding-window method (window size: 100 kb; step size: 10 kb), as described previously [69]. For SNPs, candidate outliers under strong positive selection were identified when two or more methods exhibited overlapping outlier signals (top 5% values).

For CNVs, the V_{ST} parameters were quantified to measure population differentiation in copy numbers at each CNVR between different comparisons. V_{ST} was calculated as follows: $(V_T\text{-}V_S)/V_T$, where V_T represents the total variance of copy number among all unrelated individuals, and V_S is the average variance within each population [70]. CNVRs with the highest 5% Vst values were considered selective CNVRs.

Functional enrichment analysis of candidate genes

Gene annotation for regions overlapping with selective regions was performed based on the *Bos taurus* reference genome (ARS_UCD1.2). Gene ontology (GO) analysis was conducted by Metascape [71] (Metascape, http://metascape.org). The KOBAS -i software [72] was utilized to conduct Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis through a hypergeometric test.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11328-z .

Supplementary Material 1

Supplementary Material 2

Acknowledgements
Not applicable.

Author contributions
XW carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. JP and LX participated in the bioinformatic analyses. QG, PB and LC collected the samples. XG and PY conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

Funding
This work was supported by Agricultural Science and Technology Innovation Program (25-LZIHPS-01), China Agriculture Research System of MOF and MARA (CARS-37), and Shenzhen Science and Technology Program (KCXFZ20201221173205015).

Data availability
All genomic sequencing data from this project were deposited in CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CRA008704.

Declarations

Ethics approval and consent to participate
All cattle were handled following the guidelines in accordance with the Regulations for the Administration of Affairs Concerning Experimental Animals approved by the State Council of the People's Republic of China. Informed consent was obtained from the owners of the cattle in the study. All experimental procedures in the present study were approved by the Animal Administration and Ethics Committee of Lanzhou Institute of Husbandry and Pharmaceutical Sciences of CAAS (Permit No. SYXK-2014-0002). The study is in accordance with the ARRIVE Guidelines. All operations of sample collection were allowed by the owners.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 28 May 2024 / Accepted: 5 February 2025
Published online: 28 February 2025

References

1. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, Adelson DL, Eichler EE, Elnitski L, Guigó R et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 2009;324(5926):522–8.
2. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. Proc Natl Acad Sci USA. 2008;105(33):11597–604.
3. Machugh DE, Larson G, Orlando L. Taming the past: ancient dna and the study of animal domestication. Annu Rev Anim Biosci. 2017;5:329–51.
4. Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, Hu S, Huang S, Zhang H, Zheng Z, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat Commun. 2018;9(1):2337.
5. Cai X, Chen H, Wang S, Xue K, Lei C. Polymorphisms of two Y chromosome microsatellites in Chinese cattle. Genet Sel Evol. 2006;38(5):525–34.
6. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, et al. Genetic evidence for high-altitude adaptation in Tibet. Science. 2010;329(5987):72–5.
7. Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, Cao C, Hu Q, Kim J, Larkin DM, et al. The yak genome and adaptation to life at high altitude. Nat Genet. 2012;44(8):946–9.
8. Wang MS, Li Y, Peng MS, Zhong L, Wang ZJ, Li QY, Tu XL, Dong Y, Zhu CL, Wang L, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Mol Biol Evol. 2015;32(7):1880–9.
9. Wang X, Ju Z, Jiang Q, Zhong J, Liu C, Wang J, Hoff JL, Schnabel RD, Zhao H, Gao Y, et al. Introgression, admixture, and selection facilitate genetic adaptation to high-altitude environments in cattle. Genomics. 2021;113(3):1491–503.
10. Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CK, Chen L, Ma J, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat Genet. 2013;45(12):1431–8.
11. Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, Sun YW, Zhao YX, Tu XL, Zhang M, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol. 2016;33(10):2576–92.
12. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annu Rev Genom Hum G. 2009;10:451–81.
13. Zhang Y, Hu Y, Wang X, Jiang Q, Zhao H, Wang J, Ju Z, Yang L, Gao Y, Wei X, et al. Population structure, and selection signatures underlying high-altitude adaptation inferred from genome-wide copy number variations in Chinese indigenous cattle. Front Genet. 2019;10:1404.
14. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA. 1979;76(10):5269–73.
15. Mei C, Wang H, Liao Q, Wang L, Cheng G, Wang H, Zhao C, Zhao S, Song J, Guang X, et al. Zan L. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. Mol Biol Evol. 2018;35(3):688–99.
16. Zhang Y. Animal genetic resources in China-bovines (in Chinese). Beijing: China Agriculture; 2011.
17. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010;329(5987):75–8.
18. Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Ouzhuluobu, Basang, et al. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol Biol Evol. 2011;28(2):1075–81.
19. Ge RL, Cai Q, Shen YY, San A, Ma L, Zhang Y, Yi X, Chen Y, Yang L, Huang Y, et al. Draft genome sequence of the Tibetan antelope. Nat Commun. 2013;4:1858.
20. Prabhakar NR, Semenza GL. Adaptive and maladaptive cardiorespiratory responses to continuous and intermittent hypoxia mediated by hypoxia-inducible factors 1 and 2. Physiol Rev. 2012;92(3):967–1003.
21. Wang B, Zhang YB, Zhang F, Lin H, Wang X, Wan N, Ye Z, Weng H, Zhang L, Li X, et al. On the origin of Tibetans and their genetic basis in adapting high-altitude environments. PLoS ONE. 2011;6(2):e17002.
22. Werhahn G, Senn H, Ghazali M, Karmacharya D, Sherchan AM, Joshi J, Kusi N, López-Bao JV, Rosen T, Kachel S, et al. The unique genetic adaptation of the

himalayan wolf to high-altitudes and consequences for conservation. Glob Ecol Conserv. 2018;16:e455.
23. Zhang L, Zhou Y, Wu W, Hou L, Chen H, Zuo B, Xiong Y, Yang J. Skeletal muscle-specific overexpression of PGC-1alpha induces fiber-type conversion through enhanced mitochondrial respiration and fatty acid oxidation in mice and pigs. Int J Biol Sci. 2017;13(9):1152–62.
24. Qu Y, Chen C, Xiong Y, She H, Zhang YE, Cheng Y, DuBay S, Li D, Ericson PGP, Hao Y, et al. Rapid phenotypic evolution with shallow genomic differentiation during early stages of high elevation adaptation in eurasian tree sparrows. Natl Sci Rev. 2019;7(1):113–27.
25. Scott GR, Elogio TS, Lui MA, Storz JF, Cheviron ZA. Adaptive modifications of muscle phenotype in high-altitude deer mice are associated with evolved changes in gene regulation. Mol Biol Evol. 2015;32(8):1962–76.
26. Lyu T, Zhou S, Fang J, Wang L, Shi L, Dong Y, Zhang H. Convergent genomic signatures of high-altitude adaptation among six independently evolved mammals. Animals-Basel. 2022;12(24):3572.
27. Zhang Q, Gou W, Wang X, Zhang Y, Ma J, Zhang H, Zhang Y, Zhang H. Genome resequencing identifies unique adaptations of Tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments. Genome Biol Evol. 2016;8(3):765–76.
28. Sjodin B, Russello MA. Comparative genomics reveals putative evidence for high-elevation adaptation in the American pika (ochotona princeps). G3-Genes Genom Genet. 2022;12(11):jkac241.
29. Fu TT, Sun YB, Gao W, Long CB, Yang CH, Yang XW, Zhang Y, Lan XQ, Huang S, Jin JQ, et al. The highest-elevation frog provides insights into mechanisms and evolution of defenses against high UV radiation. Proc Natl Acad Sci USA. 2022;119(46):e2082561177.
30. Yu L, Wang GD, Ruan J, Chen YB, Yang CP, Cao X, Wu H, Liu YH, Du ZL, Wang XP, et al. Genomic analysis of snub-nosed monkeys (rhinopithecus) identifies genes and processes related to high-altitude adaptation. Nat Genet. 2016;48(8):947–52.
31. Pan J, Li K, Huang W, Zhang X. MiR-137 inhibited cell proliferation and migration of vascular smooth muscle cells via targeting IGFBP-5 and modulating the mTOR/STAT3 signaling. PLoS ONE. 2017;12(10):e186245.
32. Li J, Diao S, Yang H, Cao Y, Du J, Yang D. IGFBP5 promotes angiogenic and neurogenic differentiation potential of dental pulp stem cells. Dev Growth Differ. 2019;61(9):457–65.
33. Moreno-Barriuso N, López-Malpartida AV, de Pablo F, Pichel JG. Alterations in alveolar epithelium differentiation and vasculogenesis in lungs of LIF/IGF-I double deficient embryos. Dev Dynam. 2006;235(8):2040–50.
34. Scholander PF, Hock R, Walters V, Irving L. Adaptation to cold in arctic and tropical mammals and birds in relation to body temperature, insulation, and basal metabolic rate. Biol Bull-Us. 1950;99(2):259–71.
35. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, Zhang F, Zhang L, Cui L, He W, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat Genet. 2015;47(3):217–25.
36. Xu NY, Si W, Li M, Gong M, Larivière JM, Nanaei HA, Bian PP, Jiang Y, Zhao X. Genome-wide scan for selective footprints and genes related to cold tolerance in Chantecler chickens. Zool Res. 2021;42(6):710–20.
37. Cardona A, Pagani L, Antao T, Lawson DJ, Eichstaedt CA, Yngvadottir B, Shwe MT, Wee J, Romero IG, Raj S, et al. Genome-wide analysis of cold adaptation in indigenous siberian populations. PLoS ONE. 2014;9(5):e98076.
38. Librado P, Der Sarkissian C, Ermini L, Schubert M, Jónsson H, Albrechtsen A, Fumagalli M, Yang MA, Gamba C, Seguin-Orlando A, et al. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. Proc Natl Acad Sci USA. 2015;112(50):E6889–97.
39. Lynch VJ, Bedoya-Reina OC, Ratan A, Sulak M, Drautz-Moses DI, Perry GH, Miller W, Schuster SC. Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the arctic. Cell Rep. 2015;12(2):217–28.
40. Rubin C, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464(7288):587–91.
41. Zhou L, Wu K, Zhang L, Gao L, Chen S. Liver-specific deletion of TSHR inhibits hepatic lipid accumulation in mice. Biochem Bioph Res Co. 2018;497(1):39–45.
42. Warner A, Rahman A, Solsjö P, Gottschling K, Davis B, Vennström B, Arner A, Mittag J. Inappropriate heat dissipation ignites brown fat thermogenesis in mice with a mutant thyroid hormone receptor α1. Proc Natl Acad Sci USA. 2013;110(40):16241–6.
43. Zhang J, Wu H, Ma S, Gao L, Yu C, Jing F, Zhao J. TSH promotes adiposity by inhibiting the browning of white fat. Adipocyte. 2020;9(1):264–78.
44. Zhao YD, Cai L, Mirza MK, Huang X, Geenen DL, Hofmann F, Yuan JX, Zhao YY. Protein kinase G-I deficiency induces pulmonary hypertension through rho A/Rho kinase activation. Am J Pathol. 2012;180(6):2268–75.
45. Taye M, Kim J, Yoon SH, Lee W, Hanotte O, Dessie T, Kemp S, Mwai OA, Caetano-Anolles K, Cho S, et al. Whole genome scan reveals the genetic signature of African ankole cattle breed and potential for higher quality beef. BMC Genet. 2017;18(1):11.
46. Du H, Yu J, Li Q, Zhang M. New evidence of tiger subspecies differentiation and environmental adaptation: comparison of the whole genomes of the Amur tiger and the south China tiger. Animals-Basel. 2022;12(14):1817.
47. Moazzami Z, Crawford PA, Ruan H. Ketones improve the mitochondrial coupling efficiency of brown adipocytes. FASEB J. 2022;36:R4195.
48. Gandarillas NL, Bunney TD, Josephs MB, Gierschik P, Katan M. In vitro reconstitution of activation of PLCepsilon by Ras and rho GTPases. Methods Mol Biol. 2009;462:379–89.
49. Wiener P, Robert C, Ahbara A, Salavati M, Abebe A, Kebede A, Wragg D, Friedrich J, Vasoya D, Hume DA, et al. Whole-genome sequence data suggest environmental adaptation of Ethiopian sheep populations. Genome Biol Evol. 2021;13(3):evab014.
50. Lv FH, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, Joost S, Li MH, Ajmone Marsan P. Adaptations to climate-mediated selective pressures in sheep. Mol Biol Evol. 2014;31(12):3324–43.
51. Rocha JL, Godinho R, Brito JC, Nielsen R. Life in deserts: the genetic basis of mammalian desert adaptation. Trends Ecol Evol. 2021;36(7):637–50.
52. Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, Zhang L, Abutarboush MH, Xing Y, Xie Z, et al. Camelid genomes reveal evolution and adaptation to desert environments. Nat Commun. 2014;5:5188.
53. Endres BT, Priestley JR, Palygin O, Flister MJ, Hoffman MJ, Weinberg BD, Grzybowski M, Lombard JH, Staruschenko A, Moreno C, et al. Mutation of Plekha7 attenuates salt-sensitive hypertension in the rat. Proc Natl Acad Sci USA. 2014;111(35):12817–22.
54. Ahbara AM, Musa HH, Robert C, Abebe A, Al-Jumaili AS, Kebede A, Latairish S, Agoub MO, Clark E, Hanotte O, et al. Natural adaptation and human selection of northeast African sheep genomes. Genomics. 2022;114(5):110448.
55. Luo S, Li Y, Li S, Jiang R, Deng F, Liu G, Zhang J. Expression regulation of water reabsorption genes and transcription factors in the kidneys of lepus yarkandensis. Front Physiol. 2022;13:856427.
56. Mwacharo JM, Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky BA, Rothschild MF. Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts. Sci Rep. 2017;7(1):17647.
57. Wang K, Cao Y, Rong Y, Ning Q, Jia P, Huang Y, Lan X, Dang R, Chen H, et al. A novel SNP in EIF2AK4 gene is associated with thermal tolerance traits in Chinese cattle. Animals-Basel. 2019;9(6):375.
58. Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, Rothschild MF. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity. 2016;116(3):255–64.
59. Ronaldson PT, Persidsky Y, Bendayan R. Regulation of ABC membrane transporters in glial cells: relevance to the pharmacotherapy of brain HIV-1 infection. Glia. 2008;56(16):1711–35.
60. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
62. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
63. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
64. Wang X, Zheng Z, Cai Y, Chen T, Li C, Fu W, Jiang Y. CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. Gigascience. 2017;6(12):1–12.
65. Huang Y, Li Y, Wang X, Yu J, Cai Y, Zheng Z, Li R, Zhang S, Chen N, Asadollahpour NH, et al. An atlas of CNV maps in cattle, goat and sheep. Sci China Life Sci. 2021;64(10):1747–64.
66. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

67. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
68. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.
69. Bo D, Feng Y, Bai Y, Li J, Wang Y, You Z, Shen J, Bai Y. Whole-genome resequencing reveals genetic diversity and growth trait-related genes in pinan cattle. Animals-Basel. 2024;14(15):2163.
70. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444–54.
71. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523.
72. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Res. 2021;49(W1):W317–25.

Publisher's note