Whole genome insights into genetic diversity, introgression, and adaptation of Yunnan indigenous cattle of Southwestern China

Xiwen Guan[1,2], Weixuan Xiang[2], Kaixing Qu[3], Zulfiqar Ahmed[4], Jianyong Liu[1], Ming Cai[1], Jicai Zhang[1], Ningbo Chen[2], Chuzhao Lei[2*] and Bizhi Huang[1*]

Abstract

Background Yunnan Province, located in Southwestern China, the intricate geography, variable climate, and abundant vegetation of the region have collectively contributed to shaping the distinctive germplasm characteristics observed in Yunnan indigenous cattle through prolonged domestication. The different breeds of Yunnan cattle exhibit distinct advantageous characteristics and traits, which are an important source of genetic variation because they might carry alleles that enable them to adapt to local environment and tough feeding conditions. However, a comprehensive genomic landscape of genetic resources has yet to be delineated.

Results Herein, we employed 140 whole-genome sequencing data from Yunnan indigenous cattle across eight breeds to elucidate their genetic diversity and population structure. Utilizing both uniparental and biparental markers, we elucidated the intricate genetic composition of Yunnan indigenous cattle, which is closely correlated with the geographic environment. A predominant East Asian indicine ancestry which gradually diminishes towards the north. The analysis revealed a high genetic diversity among populations and a low-to-moderate inbreeding coefficient, underscoring the rich genetic reservoir of Yunnan cattle breeds. Additionally, gene flow between Yunnan indicine and wild *Bos* species in and around Yunnan was verified, highlighting localized introgression from Yunnan Gayal as a critical factor in the successful adaptation of Yunnan indicine cattle to the local hot and humid environments.

Conclusions Our findings established the SNPs database for facilitating resource conservation and selective breeding. Moreover, these valuable insights into the genomic diversity and adaptive history of Yunnan indigenous cattle breeds contribute significantly to our understanding of their evolutionary dynamics and offer a foundation for future genetic improvement and conservation strategies.

Keywords Adaptation, Introgression, Whole-genome sequencing, Yunnan cattle

*Correspondence:
Chuzhao Lei
leichuzhao1118@126.com
Bizhi Huang
hbz@ynbp.cn
[1]Yunnan Academy of Grassland and Animal Science, Kunming, China

[2]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China
[3]Chuxiong Normal University, Chuxiong, China
[4]Department of Livestock and Poultry Production, Faculty of Veterinary and Animal Sciences, University of Poonch Rawalakot, Azad Jammu and Kashmir, Pakistan

Guan *et al. BMC Genomics*   (2025) 26:216

Page 2 of 11

Background

Domestic cattle are mainly divided into taurine (*Bos taurus*) and indicine (*Bos indicus*), originate from two primary domestication centres in the Near East and the Indus Valley [46]. Taurine cattle, known for their superior meat and milk production, can adapt to temperate and cold climates, contrast with indicine cattle which are heat-resistant and drought-resistant and adapted to tropical and subtropical climates [3]. Hybridization between indicine and taurine is one of the oldest and most effective strategies for balancing productivity, and providing new genetic resources for human and natural selection [1, 2]. China is rich in bovine species resources, and there are 55 native cattle breeds exhibiting distinct phenotypes [7] which have been shown to have five types of ancestry: Eurasian taurine, European taurine and East Asian taurine in northern China, East Asian indicine and South Asian cattle in southern China [4].

Yunnan Province serves as a pivotal region for the entrance of indicine cattle into East Asia through the inland routes [4], acting as the hybrid region of taurine and indicine cattle. The migration of Indian cattle through Yunnan into China has significantly enriched the genetic diversity of Yunnan cattle [5]. Yunnan boasts a rich diversity of cattle breeds, providing essential resources such as hides, meat, and milk to residents and communities [6]. These breeds are not only pivotal for transportation and ploughing in remote farms but also exhibit admirable traits such as superior foraging behaviour, heat and humidity resistance, and disease resistance [7].

Yunnan Province officially recognizes six indigenous cattle breeds: Wenshan cattle, Dengchuan cattle, Dianzhong cattle, Dehong Humped cattle, Zhaotong cattle, and Diqing cattle, and locally distinctive cattle populations: Jiangcheng cattle, and Lincang Humped cattle. These eight cattle breeds each possess unique characteristics but face varying degrees of genetic degeneration and survival challenges. Wenshan cattle exhibit strong adaptability but poor hindquarter development, Diqing cattle are cold-resistant but have lower reproductive performance, Dianzhong cattle are compact with tender meat but suffer from inbreeding issues, Dengchuan cattle have excellent milk quality but are endangered and require protection, Dehong Humped cattle thrive in tropical climates but lack sufficient breeding efforts, Zhaotong cattle are cold-resistant but exhibit poor hindquarter development, and Jiangcheng cattle, characterized by their small size and muscular tenderness with a historical trade significance [7]. Moreover, Yunnan represents a reservoir of bovine biodiversity, harboring not only cattle but also other *Bos* species like gayal (*Bos frontalis*), banteng (*Bos javanicus*), yak (*Bos grunniens*), buffalo (*Bubalus bubalis*), and gaur (*Bos gaurus*), the genetic introgression events between wild species and domesticated cattle have enriched the genetic diversity [7, 8].

Previous assessments of indigenous Yunnan cattle have primarily focused on mitochondrial diversity, kinship with introduced beef breeds, and genomic diversity and hybridization patterns based on SNP chip data [6, 10]. With the development of sequencing technology and the reduction of sequencing costs, whole genome sequencing (WGS) technology has been responsible for several milestones in exploring population structure, demographic history and economic and environmental adaptation traits of these crucial livestock genetic resources [11, 12]. Given the complex geographic terrain of Yunnan and the limited exploration of whole-genome information for domestic cattle and other bovine subfamily species, in this study, we analyze the genomes of 11 Jiangcheng cattle, 10 Dianzhong cattle, one Diqing cattle, and 13 Wenshan cattle (Table S1). Additionally, we collected whole-genome data for 105 Yunnan native cattle (Table S1), fifteen Yunnan gayal, three Zhongdian yak and 111 Bovinae to serve as a reference group (Table S2). The aim of this study was to explore the unique genomic characteristics and phylogeographic patterns of the diversity of Yunnan cattle using the largest Yunnan cattle genome dataset available to date. Our research identified the genome diversity, population structure, and extensive introgression of Yunnan gayal that enable Yunnan cattle to adapt local extreme environments. The insights gained from this research will provide valuable genetic information, crucial for improving the breeding and genetics of Yunnan cattle, aiming at their sustainable development and conservation.

Methods

Sample collection and genome sequencing

Ear tissue samples were randomly collected from Dianzhong cattle ($n = 10$), Wenshan cattle ($n = 13$), Jiangcheng ($n = 11$), Diqing ($n = 1$) of Yunnan province, China. We used a standard phenol/chloroform-based protocol to extract the ed genomic DNA of the ear tissue samples. The DNA library was constructed for each sample (500 bp insert size). Sequencing via Illumina NovaSeq 6000 with 2×150 bp model at Novogene Bioinformatics Institute, Beijing, China, and 150 bp paired-end sequence data were generated (Table S1). Furthermore, the genome data of 234 bovines were downloaded from the NCBI database, including 105 Yunnan native cattle, 16 East Asian indicine, 11 South Asian indicine, 15 East Asian taurine, 19 Eurasian taurine, 25 European taurine, 3 Zhongdian yak, 15 Yunnan gayal, 10 Bangladeshi gayal, 3 Indian gayals, 2 Bisons, 2 Wisent, 2 banteng, 4 gaur and 2 buffalo (Table S2).

Reads mapping and SNP identification

All processed sequencing reads were aligned to the Bos taurus reference genome ARS-UCD1.2 utilizing BWA-MEM [13], achieving a mean mapping rate of 99.71%. Potential duplicate reads were identified and removed using Picard tools (http://broadinstitute.github.io/picard) with the parameter REMOVE_DUPLICATES set to true. Subsequent SNP detection was performed using the Genome Analysis Toolkit (GATK) [48], adhering to the following stringent criteria: quality by depth (QD) less than 2; root mean square (RMS) mapping quality greater than 40.0; Phred-scaled *P* value from Fisher's exact test for strand bias below 60.0; Z score from the Wilcoxon rank sum test for alternate versus reference read mapping qualities (MQs) exceeding – 12.5; Z score from the Wilcoxon rank sum test for alternate versus reference read position bias greater than – 8.0; and mean sequencing depth (across all individuals) < 1/3× and > 3×. Finally, ANNOVAR [14] was utilized to annotate the distribution of SNPs across various genomic regions specific to each breed.

Population structure analyses in whole genome sequencing data

SNPs exhibiting high levels of pairwise linkage disequilibrium (LD) were pruned using PLINK [15], with the parameters set to (--indep-pairwise 50 5 0.2). Following this, population structure analyses were conducted on a dataset comprising 160 Yunnan native cattle and five additional "core" cattle populations. For these analyses, we utilized multiple approaches: the neighbor-joining (NJ) tree, principal component analysis (PCA), and ADMIXTURE. The NJ tree was constructed using PLINK and subsequently visualized with the aid of MEGA [16] and FigTree (http://tree.bio.ed.ac.uk/software/figtree/). PCA was carried out employing SmartPCA, included within the EIGENSOFT [51], to elucidate the principal components of the SNP data. The population genetic structure was inferred using ADMIXTURE [17], considering 2 to 6 clusters (*K*).

Genomic diversity

We conducted a comparative analysis of nucleotide diversity among eight Yunnan cattle breeds and five "core" cattle populations. Nucleotide diversity for each breed was assessed in 50 kb windows with 50 kb steps, utilizing VCFtools [18]. Linkage disequilibrium (LD) decay was estimated based on the physical distance between pairwise SNPs using PopLDdecay software [19], employing default parameters. The length and number of runs of homozygosity (ROH) for each individual were computed using VCFtools with the following parameters: -homozyg-density 50 -homozyg-window-het 3 -homozyg-window-missing 5. ROHs for each breed were categorized

into four length classes: 0.5–1 Mb, 1–2 Mb, 2–4 Mb, and greater than 4 Mb. F_{ROH} was calculated as the proportion of the genome in ROH overall the length of the autosomal genome: F_{ROH} = L(ROH)/L(Autosomes), where L(ROH) is the sum of all ROH of a sample, and L(Autosomes) is the total length of the autosomal genome covered by analysed SNPs.

Estimates of the effective population size

An MSMC was used to infer effective population sizes (Ne) considering 2 samples. Autosomal SNPs of each sample were identified using GATK. After removing variant outliers (with extremely low or high coverage), all sites were phased using BEAGLE [20]. The time scale is calculated using an average generation time of 6 years (g = 6) and a mutation rate of $\mu g = 1.26 \times 10^{-8}$.

Whole-mitochondrial genome phylogenies and Y-chromosome

A total of 140 mitochondrial genomes were reconstructed from whole-genome resequencing data, supplemented by 12 additional complete mtDNA sequences retrieved from GenBank. Phylogenetic analysis was conducted on the final sequence alignment using RAxML [21] with the following parameters: -f a -x 123 -p 23 -# 100 -k -m 132 GTRGAMMA. The resultant phylogenetic tree topology was visualized using FigTree. A median-joining network was generated using NETWORK software [22]. For Y-chromosome analysis, we focused on the X-degenerate region containing single-copy genes within the male-specific region of the Btau_5.0.1 Y-chromosome reference sequence (GCF_000003205.7), involving 106 Yunnan cattle and 5 reference individuals. A total of 608 SNPs were extracted for analysis. Haplogroup trees were constructed from FASTA-formatted sequences using maximum likelihood (ML) methods.

Introgression analysis

To evaluate gene flow direction, *D*-statistics were computed using ADMIXTOOLS [23]. Under the null hypothesis of no gene flow, the *D*-statistic is expected to be zero. The *D*-statistic was calculated based on the tree topology [[[Pop A (W), Pop B (X)], Pop C (Y)], outgroup (Z)], where the buffalo serves as the outgroup and South Asian Indicine (SAI) cattle as Pop A. Six tree topologies were used for introgression analysis: *D* (SAI cattle, Yunnan cattle, banteng, buffalo), *D* (SAI cattle, Yunnan cattle, gaur, buffalo), *D* (SAI cattle, Yunnan cattle, Zhongdian yak, buffalo), *D* (SAI cattle, Yunnan cattle, Yunnan gayal, buffalo), *D* (SAI cattle, Yunnan cattle, Bangladeshi gayal, buffalo), and *D* (SAI cattle, Yunnan cattle, Indian gayal, buffalo). Negative *D*-statistics suggest gene flow from Y to X, while positive *D*-statistics indicate gene flow from Y to W. Only *D*-statistics with a *Z*-score exceeding ± 3

were considered significant. Additionally, we employed the --supervised parameter in ADMIXTURE software to calculate the introgression proportion of Yunnan indigenous cattle from outgroup (wild species).

For regions with confirmed gene flow in Yunnan cattle, further analysis was conducted. The U20 statistic (SAI, Yunnan cattle, Yunnan gayal $_{(1\%, 20\%, and 100\%)}$) [24] was employed to identify genomic regions associated with adaptive introgression, defined by the presence of alleles fixed in Yunnan gayal but occurring at less than 1% frequency in SAI cattle or greater than 20% in Yunnan cattle. Phylogenetic analysis was performed on the introgressed region BTA5:100,470,001−100,570,000. Functional annotation of genes within this region was conducted using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) terms, utilizing KOBAS (http://kobas.cbi.pku.edu.cn/) [25]. Results were deemed significantly enriched at a corrected *P*-value threshold of <0.05.

Results

Sequence read depth and genomic landscapes of analyzed dataset

The whole genomes of 35 samples were sequenced to an average of ∼10.01× depth coverage and jointly genotyped with 229 publicly available genomes, including the published Yunnan cattle, five "core" cattle populations, and wild *Bos* species (Table S2). Mapping with the taurine reference genome (ARS-UCD1.2) generated reads with an average alignment rate of 99.40% (minimum: 86.74%, maximum: 99.89%). In total, ∼89 million SNPs were finally retained, and each breed-specific SNPs were identified. Functional annotation of the polymorphic sites revealed that the vast majority of SNPs were present in either intergenic regions or intronic regions (Table S3). We counted the specific SNPs of eight breeds; the largest number (37,176, 443) and specific SNPs (2,824,259) were found in Wenshan cattle (Table S4). The difference in the number of SNPs and specific SNPs could be partly explained by the different sample sizes per breed and partly by breed differences.

Population genetic structure and ancestral relationships

In our investigation into the population genetic structure and ancestral relationships of Yunnan cattle, we utilized genetic clustering analysis on a dataset comprising 205 individuals, including native cattle from Yunnan alongside five "core" cattle populations. Our findings highlighted a predominant admixture of taurine and indicine ancestries across Yunnan, with a notable gradient where the influence of indicine ancestry is most pronounced in the southern regions and diminishes towards the north. The cattle breeds were divided into taurine or indicine ancestry at $K = 2$, while at $K = 3$, Dehong cattle displayed

a nearly uniform ancestral composition, sharing ancestry with South Asian indicine. This pattern became increasingly discernible at $K = 4$, demonstrating the complex genetic makeup of these populations (Fig. 1B). Further differentiation among breeds was evident through the construction of a neighbor-joining (NJ) tree, which, with few exceptions, grouped each breed into its own clade, suggesting distinct genetic lineages despite regional hybridization (Fig. 1C). PCA reinforced these findings, illustrating the genetic divide between indicine and taurine cattle along the first principal component and further separating East Asian taurine cattle from European breeds along the second principal component. A distinct partition between South Asian indicine and East Asian indicine breeds was also observed, placing Yunnan cattle in an intermediary position along PC1, with most breeds leaning towards indicine ancestry except for Diqing breed. Notably, the Wenshan breed emerged as genetically distinct from other Yunnan groups, aligning more closely with East Asian indicine (Fig. 1D). Overall, the population genetics analysis revealed an intermediate genetic positioning of most Yunnan cattle between East Asian indicine and South Asian indicine ancestries, with a few breeds showing a blend of East Asian taurine and indicine lineages. This gradient of ancestral influence from north to south within Yunnan mirrors the broader distribution trends observed across China [4].

Genome diversity and demographic history

The nucleotide diversity analysis showed that indicine origin cattle, including East Asian indicine, Yunnan cattle and South Asian indicine, had significantly higher nucleotide diversity than taurine cattle (Fig. 2A). The study further explored genome-wide linkage disequilibrium (LD) across all breeds, observing a rapid decrease in LD within the first 50 kb. Notably, the average LD in Yunnan cattle was found to be lower than that in European taurine cattle (Angus) and East Asian taurine (Hanwoo), indicating a more diverse genetic background among Yunnan breeds (Fig. 2B). Specifically, among Yunnan breeds, Diqing and Dengchuan cattle showed the highest inbreeding coefficients (Fig. 2C). When examining ROH pattern of Yunnan cattle, we divided the length of ROH into four size classes: 0.5–1 Mb, 1–2 Mb, 2–4 Mb, > 4 Mb (Fig. 2D). The presence of long ROH is the result of consanguineous mating, whereas shorter ROH reflect distant ancestral influences [26]. It was evident that commercial breeds have more medium (2–4 Mb) and long ROH (> 4 Mb) compared to indigenous cattle.

Later, employing the MSMC method, we estimated the effective population sizes (N_e) of eight Yunnan cattle breeds and representative breeds. Remarkably, a significant decline in effective population size was observed for these breeds around 20,000 to 30,000 years ago,
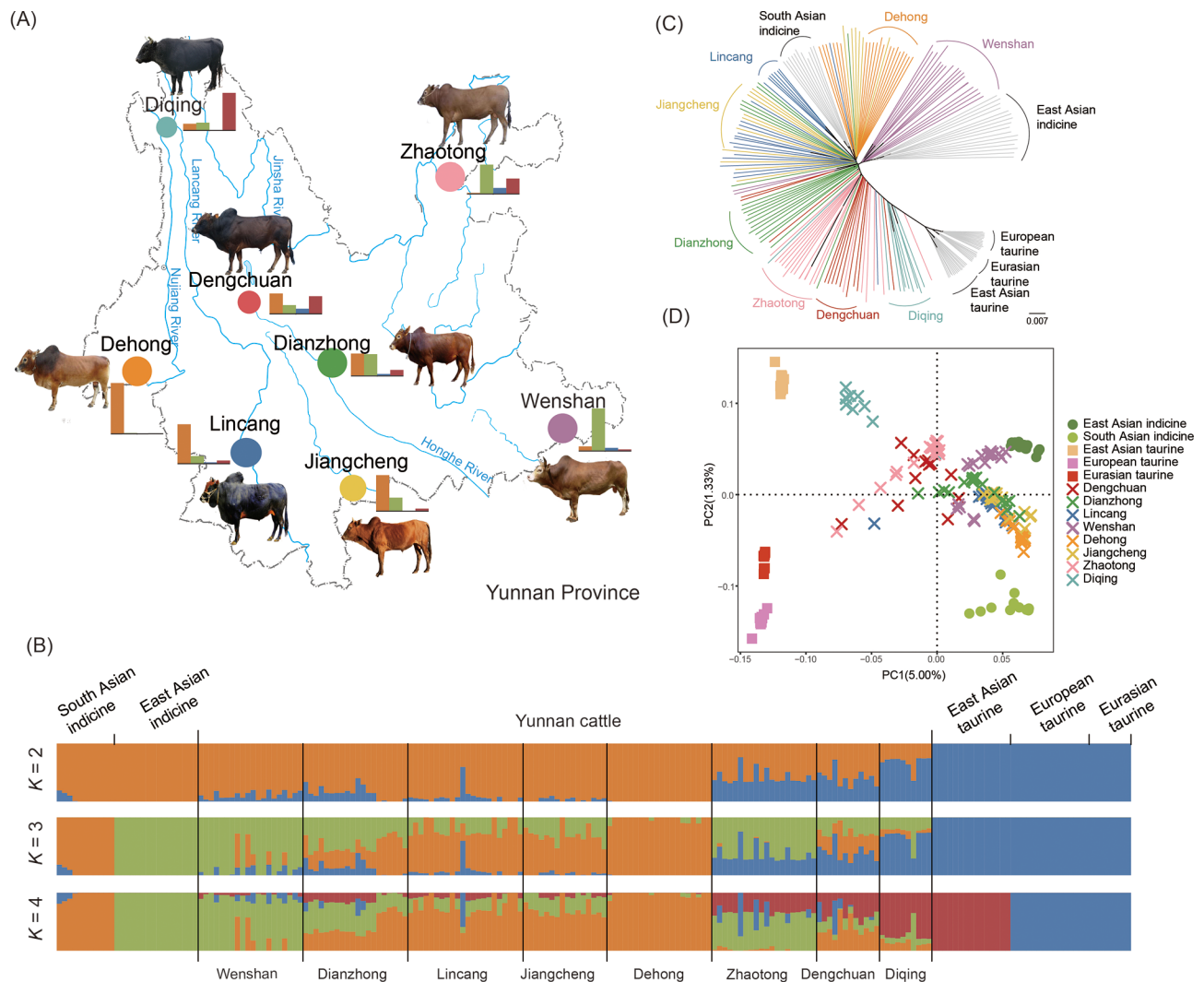
Fig. 1 Population structure of Yunnan cattle in comparison to several possible ancestral breeds. (**A**) Geographical locations of 140 Yunnan indigenous cattle breeds included in this study. The bar chart represents the ancestry proportions of each breed based on ADMIXTURE results. Orange represents South Asian indicine cattle, green represents East Asian indicine cattle, red represents East Asian taurine cattle, and orange represents European taurine cattle. (**B**) Model-based clustering of cattle breeds using ADMIXTURE with *K*=2, *K*=3 and *K*=4. (**C**) Neighbor-joining tree of the relationships between Yunnan cattle and other cattle. (**D**) Principal component analysis

coinciding with the Last Glacial Maximum-a period characterized by substantial cooling (Fig. 2E), predating cattle domestication. We also observed a decline of N_e during 7–9 kya consistent with the onset of domestication. Apart from Diqing cattle, we observed that Yunnan cattle exhibit a similar effective population size pattern as East Asian indicine (Wenshan cattle). This may be related to Diqing cattle inhabiting high-altitude areas adjacent to the Qinghai-Tibet Plateau.

Furthermore, genetic distance measurements among these breeds were assessed using an F_{ST} matrix, revealing variability ranging from 0.01 to 0.2. Diqing cattle show notably higher genetic differentiation from other groups with most breeds exhibiting subtle differences (Fig. 2F).

Uniparental phylogenies

In our exploration of uniparental phylogenies, the study explored the origins of paternity and maternity among Yunnan indigenous cattle through the analysis of Y-chromosome and mitochondrial DNA (mtDNA) (Fig. S1). By constructing a mitochondrial DNA haplotype network based on whole-genome sequences; we discovered two major maternal lineages among these cattle: one associated with taurine cattle and the other with indicine cattle (Fig. 3A and Table S6). Remarkably, except for the Diqing cattle, which were traced back to pure taurine maternal origins, the remaining seven breeds of Yunnan indigenous cattle were found to originate from a mix of taurine and indicine maternal lineages. Notably, one mtDNA
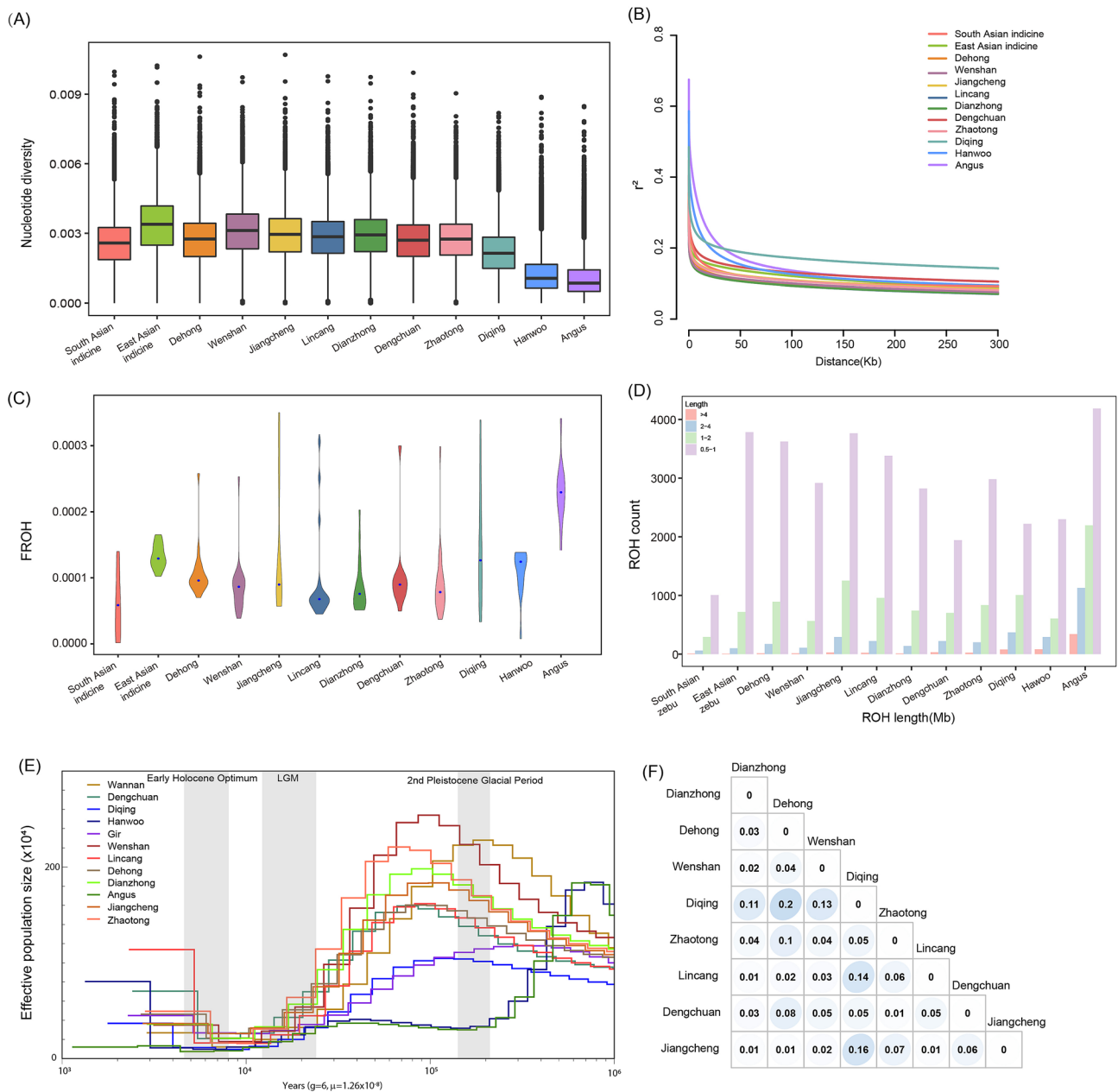
Fig. 2 Characterization of genomes among Yunnan cattle and reference populations. (**A**) Box plots of nucleotide diversity in 50-kb sliding windows with 20-kb steps. (**B**) Decay of linkage disequilibrium on cattle autosomes estimated from each breed (**C**) Inbreeding coefficient for each breed. (**D**) Run of homozygosity (ROH) distribution in and among breeds. (**E**) Population size history inference of 8 Yunnan cattle breads and 4 reference cattle breeds using MSMC. (**F**) Mean F_{ST} values between pairwise breeds

sequence from Dengchuan cattle was closely aligned with the yak haplotype.

Further analysis using the Y-chromosome haplotype network of 106 male cattle revealed the division of Yunnan cattle into two major paternal lineages: those related to taurine cattle (Y2a and Y2b haplotypes) and those pertaining indicine cattle (Y3a and Y3b haplotypes) (Fig. S2). The study identified that the majority of Yunnan indigenous cattle primarily belong to the Y3a and Y3b haplotypes, indicative of a predominantly indicine paternal

origin. Jiangcheng cattle, Lincang Humpe cattle, and Dehong cattle have a sole paternal lineage of indicine origin, resulting in stable genetic ancestry and high homozygosity (Fig. 3B and Table S7).

The introgressed event from wild bovine species

We then assessed whether the local environmental adaptation of Yunnan cattle may have involved introgression from other bovine species living in Yunnan Province and neighboring countries. Our analysis identified significant
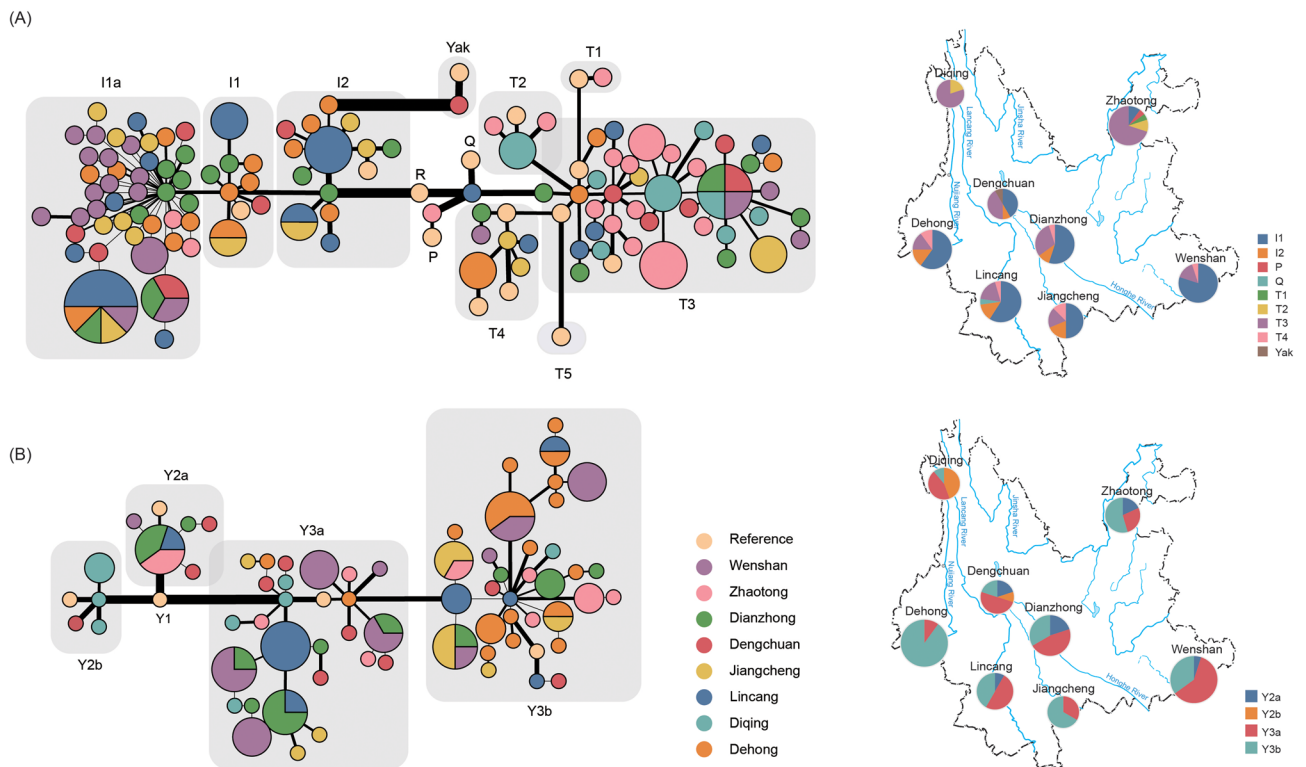
Guan *et al. BMC Genomics*   (2025) 26:216

Page 7 of 11

Fig. 3 Mitogenome phylogenies and Y-chromosome. Different haplogroups were labelled in each gray shadow. Each slice in the circles represents an individual. The width of the edges is proportional to the number of pairwise differences between the joined haplotypes. (**A**) Median-joining (MJ) network of mitogenome from Yunnan cattle. (**B**) MJ network of Y-chromosome haplotypes using 608 SNPs

introgressive hybridization events from banteng, gaur, Zhongdian yak, Yunnan gayal, Bangladeshi gayal, and Indian gayal into eight Yunnan cattle breeds, as evidenced by the *D*-statistic (Table S8). The results highlighted notable introgression from banteng, gaur, and gayal into several Yunnan breeds, including those from Zhaotong, Dehong, Lincang, Dianzhong, Jiangcheng, and Wenshan. Additionally, genetic flow between Zhongdian yak and cattle breeds from Diqing, Dengchuan, and Zhaotong was observed. The patterns of introgression from Yunnan gayal and Bangladeshi gayal were found to be mainly concentrated in the cattle breeds from Dehong, Lincang, Dianzhong, Jiangcheng, and Wenshan (Fig. 4A). Furthermore, we used the --supervised parameter in ADMIXTURE to estimate the introgression events for each individual of Yunnan cattle, and the results were close to the *D*-statistic (Fig. S3).

Using a $P < 0.05$ cutoff for the frequencies of Yunnan gayal derived alleles in the Yunnan cattle genomes (U20 South Asian indicine (SAI), Yunnan cattle, Yunnan gayal (1%, 20%, 100%)), 511 candidate regions were shortlisted (Fig. 4B and Table S9). These genes were enriched in KEGG pathways and GO terms (Corrected *P* value < 0.05) of Inflammatory mediator regulation of TRP channels, Human cytomegalovirus infection, Thermogenesis, interleukin-1 receptor binding, inflammatory response to antigenic stimulus and Thyroid hormone synthesis signaling pathway (Fig. 4C and Table S10). This enrichment suggests that the introgressed genomic regions may play significant roles in the environmental adaptation and physiological regulation of Yunnan cattle, underscoring the evolutionary impact of introgression on these indigenous breeds.

Discussion

As the beef industry continues to intensify, the survival space for indigenous cattle breeds, particularly those in Yunnan, faces significant challenges, heightening the difficulty of their conservation [6]. In response to these challenges, we conducted a comprehensive landscape genomic analysis utilizing the most extensive dataset of whole-genome sequence variations available for Yunnan cattle. Such population genetic studies are crucial for elucidating ancestral lineages and evolutionary origins in livestock.

The migration of taurine cattle from West Asia, the domestication center, eastward to northern China occurred approximately 5,000 to 4,000 years ago. Indicine cattle likely entered East Asia between 3,500 and 2,500 years before the present (YBP) [27]. The spread of indicine, on the other hand, originated from the Indian subcontinent, initially introduced to southwestern China and gradually expanding to the northern regions. Yunnan
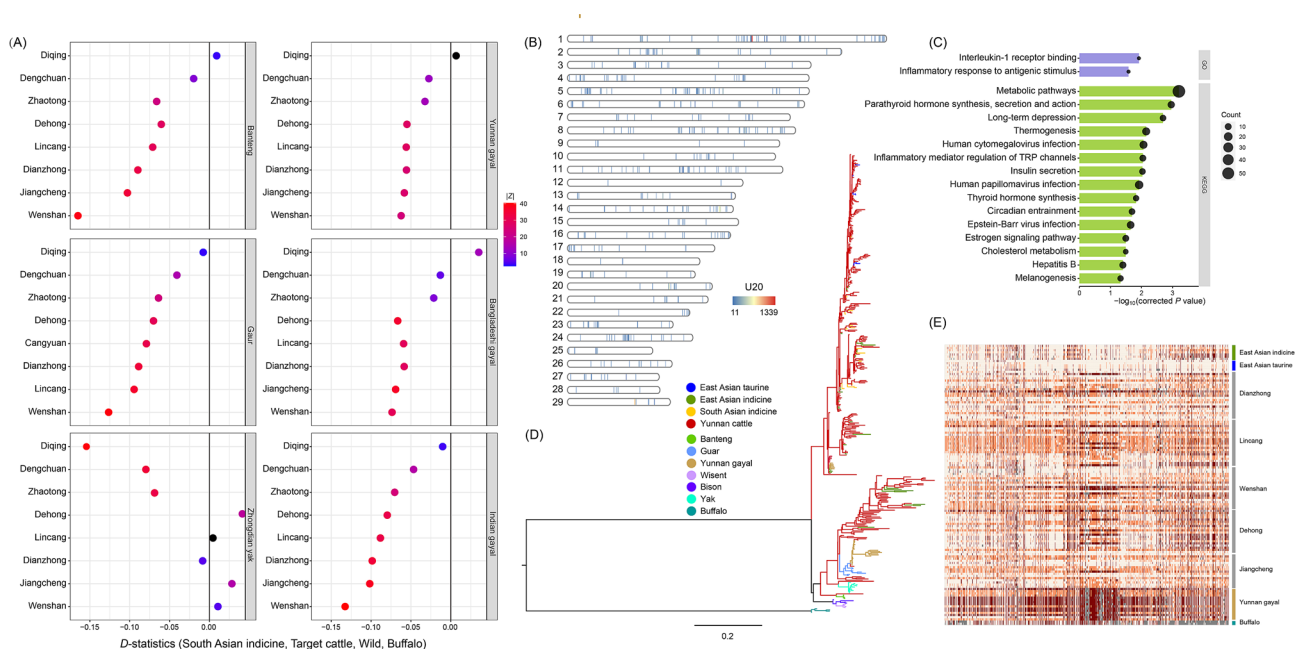
Fig. 4 Genome-wide introgression from Yunnan wild *Bos* species into Yunnan domestic cattle. (**A**) Allele sharing between Yunnan cattle and wild species. The blank line represents *D*=0, the black circle represents |*Z*| < 3. (**B**) Map of the lengths and distributions of putatively adaptative introgressed segments in the Yunnan cattle autosomes according to the results of the U20 statistic. (**C**) The KEGG and GO pathways from the enrichment analysis of the introgressed segments from Yunnan gayal into Yunnan cattle. (**D**) Phylogenetic trees were constructed using the haplotype sequences from the BTA5:100,470,001−100,570,000 region. (**E**) SNPs with MAF > 0.05 were used to build haplotype patterns from the BTA5:100,470,001−100,570,000 region

may have served as the first point of entry for South Asian indicine into China [4, 28]. However, the rugged landscape of Yunnan, intersected by three major rivers (Nujiang, Lancang, and Jinsha) (Fig. 1A), acted as a geographical barrier, impeding hybridization between South Asian indicine and East Asian indicine.

Our autosomal data analysis revealed a distinct genetic distribution pattern among eight Yunnan cattle breeds, showing a gradual decrease in taurine ancestry from north to south and a reciprocal decrease in indicine ancestry from south to north. Notably, Dehong cattle emerged through centuries of acclimatization and selective breeding from "Gala" cattle introduced from Myanmar and Nujiang River effectively isolated the Dehong region, resulting in Dehong cattle predominantly possessing South Asian indicine ancestry. Conversely, Wenshan cattle, located closer to Guangxi and Hainan province (Southern China), exhibit a relatively pure East Asian indicine ancestry.

Furthermore, our analysis of mitochondrial DNA (mtDNA) and Y-chromosome haplotypes unveiled the presence of yak mtDNA in Dengchuan cattle, similar to the finding previously reported only in Diqing cattle [29]. Within the Yunnan cattle population, one Zhaotong cattle was classified under haplogroup P and one Lincang Humped cattle under haplogroup Q, suggesting ancient bovine influences possibly retained through geographical isolation over extended breeding periods [30].

Intriguingly, the mtDNA of Dehong and Wenshan cattle possess haplogroups T and I, indicating that these breeds have been influenced by taurine. In terms of paternal lineage, Dehong cattle, Wenshan cattle, and Jiangcheng cattle have a purely indicine origin. Diqing cattle show a significant East Asian taurine influence but lack any European taurine ancestry, this distribution pattern is consistent with Tibetan cattle [47]. The lineage distribution of Yunnan cattle corresponds to the geographic distribution patterns observed among Chinese cattle breeds, thus mirroring the migration of taurine cattle into Northern China and the indicine's subsequent spread from the Southwest [5, 31].

Livestock's genetic diversity effectively reflects various historical events such as bottlenecks, genetic drift, natural selection, and domestication. Compared to commercial breeds, Yunnan indigenous cattle exhibit higher levels of nucleotide diversity, indicating a rich genetic diversity. Additionally, we conducted runs of homozygosity (ROH) analysis on each individual. The relatively low total length and number of ROH segments in the eight Yunnan indigenous cattle breeds suggest extensive hybridization within these populations. Diqing cattle, inhabiting secluded high-altitude areas, exhibit higher levels of inbreeding, while Dengchuan cattle, the only dairy breed nationwide, have undergone extensive selective breeding, resulting in higher inbreeding coefficients. Given that only Diqing cattle possess a substantial

amount of taurine ancestry, genetic differentiation among Yunnan indigenous cattle breeds, excluding Diqing cattle, remains inconspicuous, as indicated by the result of mean F_{ST} values. Moreover, when the estimated genetic distance (mean F_{ST} values) between two populations exceeds or approaches 0.05, subsequent population selection analysis can be conducted even with fewer than 20 individuals per population [32]. Therefore, the F_{ST} values between population pairs provide theoretical references for future studies.

Compared to commercial northern breeds, indigenous cattle exhibit genetic advantages in disease resistance, heat tolerance, and adaptation to local environmental conditions. The migratory patterns and climatic influences have facilitated the introgression of Indicine cattle with regional species such as gayal and banteng [33], aiding in the domestication process. Conversely, regional species like gaur, gayal, and banteng have contributed to the genetic diversity of domestic Indicine cattle, enhancing their adaptability to local environments [4, 9]. Our study confirmed that introgression events involving closely related species from Yunnan and neighboring regions have significantly augmented the genetic diversity of Yunnan indigenous cattle.

Yunnan gayal, a semi-domesticated animal, is mainly distributed in the narrow valley of the Dulong river in western Yunnan, China, and the adjacent area, also named 'Dulong' cattle [34]. As in our results of *D*-statistic, we identified several regions that have been introgressed from Yunnan gayal into Yunnan domestic cattle, which may have facilitated their adaptation to the humid tropics and subsequent rapid dispersal. For example, the top region on BTA1 (66.69–67.07 Mb) harbored *HSP-BAP1* gene relevant to heat stress, which encodes a protein that binds to one of the small heat shock proteins, specifically hsp27 which belongs to the small molecular weight heat shock protein (HSP) family (12–43 kDa) and *HSP27* was characterized in response to heat shock [35] as a protein chaperone that facilitates the proper refolding of damaged proteins [36, 37]. Additionally, the region on BTA5: 100,470,001–100,570,000 exhibits discernible introgression patterns in Yunnan cattle (Fig. 4D and E), which contained the *KLRB1* gene. Killer cell lectin-like receptor subfamily B member 1 (*KLRB1*), which encodes the inhibitory T cell receptor called *CD161*, plays an inhibitory role on natural killer (NK) cells cytotoxicity [38]. *CD47* acts as an immune checkpoint receptor, safeguarding cells from phagocytic clearance by immune cells, including macrophages, via interaction with SIRPα on the cell surface [39, 40]. Moreover, superoxide dismutase 1 (*SOD1*), degraded the hydrogen peroxide to oxygen and water for protecting the cells from oxidative damage, which plays an important role in heat tolerance in African indicine cattle [41, 42]. These genes are pivotal in immune response activation and adaptation to environmental stressors such as heat, underscoring their significance as candidate genes influencing tropical adaptation.

Previous population analyses based on SNPs have confirmed the genetic diversity of African, Asian, and European cattle [4, 11, 43, 49]. Chinese cattle have been shown to have three types of ancestry: Eurasian taurine and East Asian taurine in northern China and Chinese indicine in southern China [4], which are inseparable from the complex genetic background and domestication history of Chinese cattle. Following their contact, a north-to-south taurine-to-indicine cline of cattle was established. Meanwhile, the hybrid breeds in China have different genetic backgrounds and varying levels of genetic contributions, and selection played a role in shaping the taurine × indicine admixture proportion in hybrid cattle [12]. Compared to other provinces in China, Yunnan province was the first entry point for Indian indicine cattle into China. Additionally, dense vegetation in the region provides abundant resources for wild *Bos* species. The complex geographical environment and long-term artificial selection have contributed to the unique cattle breeds found in Yunnan. We conducted a comprehensive landscape genomic analysis of whole genome sequence variations in the largest dataset available to date for Yunnan cattle. Whether through natural selection or human-mediated processes, introgressive hybridization has enriched the genetic diversity of Yunnan cattle, thereby presenting valuable opportunities for native cattle breeding efforts. With the advancement of sequencing technologies in the future, we may explore other types of variations [44, 45, 50] thereby fully utilizing the abundant genetic resources of Yunnan indigenous cattle.

Conclusion

To sum up, by analyzing the whole-genome data of Yunnan cattle, we evaluated the genetic diversity of Yunnan cattle and multidimensionally explored the population structure of Yunnan cattle. In addition, we tested for excessive allele sharing between Yunnan indicine and wild *Bos* species that inhabited Yunnan Province or peripheral area. Our research ultimately published the catalog of genetic variants in Yunnan cattle, providing a basis for genetic breeding and resource protection in Yunnan cattle. Moreover, we identified introgression events from wild *Bos* species, particularly Yunnan gayal, contributing to the adaptation of Yunnan cattle to local environmental challenges. The establishment of a comprehensive SNP database serves as a valuable resource for development of reasonable breeding strategies and resource conservation efforts of Yunnan indigenous cattle, thereby ensuring their continued contribution to the economy and cultural heritage of the region.

Guan *et al. BMC Genomics*   (2025) 26:216

Page 10 of 11

Abbreviations

SNP	Single nucleotide polymorphism
WGS	Whole-genome sequencing
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
NJ	Neighbor-joining
PCA	Principle component analysis
ROH	Runs of homozygosity
LD	Linkage disequilibrium
mtDNA	mitochondrial DNA

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11033-3.

Supplementary Material 1

Supplementary Material 2

Author contributions

XG, CL and BH conceived and designed the experiments. XG and WX performed the statistical analysis. NC and CL provided technical assistance. KQ, JL, MC, JZ, and BH contributed to the sample collections. XG and ZA drafted the manuscript. All authors read and approved the final manuscript.

Data availability

All raw sequencing data have been submitted to the NCBI SRA under the BioProject accession number PRJNA1108623.

Declarations

Ethics approval and consent to participate

The experimental procedures were approved by the Experimental Animal Management Committee (EAMC) of Northwest A&F University (2011-31,101,684) and complied with the National Standard of Laboratory Animals Guidelines for Ethical Review of Animal Welfare (GB/T 35892 – 2018) and Guide for the Care and Use of Laboratory Animals: Eighth edition.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P. Evidence for two independent domestications of cattle. Proc Natl Acad Sci U S A. 1994;91(7):2757–61.
2. Xia X, Qu K, Wang Y, Sinding MS, Wang F, Hanif Q, Ahmed Z, Lenstra JA, Han J, Lei C, et al. Global dispersal and adaptive evolution of domestic cattle: a genomic perspective. Stress Biol. 2023;3(1):8.
3. MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG. Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (Bos taurus and Bos indicus). GeneticsGenetics 1997;146(3):1071–1086.
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. Jia S, Chen H, Zhang G, Wang Z, Lei C, Yao R, Han X. Genetic variation of mitochondrial D-loop region and evolution analysis in some Chinese cattle breeds. J Genet Genomics. 2007;34(6):510–8.
6. Yu Y, Lian LS, Wen JK, Shi XW, Zhu FX, Nie L, Zhang YP. Genetic diversity and relationship of Yunnan native cattle breeds and introduced beef cattle breeds. Biochem Genet. 2004;42(1–2):1–9.
7. Zhang Y. Animal genetic resources in China-bovines (in Chinese). Beijing: China Agriculture; 2011.
8. Gou X, Wang Y, Yang S, Deng W, Mao H. Genetic diversity and origin of Gayal and cattle in Yunnan revealed by mtDNA control region and SRY gene sequence variation. J Anim Breed Genet. 2010;127(2):154–60.
9. Wu DD, Ding XD, Wang S, Wojcik JM, Zhang Y, Tokarska M, Li Y, Wang MS, Faruque O, Nielsen R, et al. Pervasive introgression facilitated domestication and adaptation in the Bos species complex. Nat Ecol Evol. 2018;2(7):1139–45.
10. Li R, Li C, Chen H, Liu X, Xiao H, Chen S. Genomic diversity and admixture patterns among six Chinese indigenous cattle breeds in Yunnan. Asian-Australas J Anim Sci. 2019;32(8):1069–76.
11. Kim J, Hanotte O, Mwai OA, Dessie T, Bashir S, Diallo B, Agaba M, Kim K, Kwak W, Sung S, et al. The genome landscape of indigenous African cattle. Genome Biol. 2017;18(1):34.
12. Lyu Y, Ren Y, Qu K, Quji S, Zhuzha B, Lei C, Chen N. Local ancestry and selection in admixed Sanjiang cattle. Stress Biol. 2023;3(1):30.
13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
14. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
15. 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.
16. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing platforms. Mol Biol Evol. 2018;35(6):1547–9.
17. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011;12:246.
18. 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):2156–2158.
19. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35(10):1786–8.
20. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81(5):1084–97.
21. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90.
22. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16(1):37–48.
23. Harney E, Patterson N, Reich D, Wakeley J. Assessing the performance of qpAdm: a statistical tool for studying population admixture. Genetics 2021;217(4).
24. Racimo F, Marnetto D, Huerta-Sanchez E. Signatures of archaic adaptive introgression in Present-Day Human populations. Mol Biol Evol. 2017;34(2):296–317.
25. 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.

26. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. BMC GENET. 2012;13:70.
27. Chen S, Lin BZ, Baig M, Mitra B, Lopes RJ, Santos AM, Magee DA, Azevedo M, Tarroso P, Sasazaki S, et al. Zebu cattle are an exclusive legacy of the South Asia neolithic. Mol Biol Evol. 2010;27(1):1–6.
28. Felius M, Beerling ML, Buchanan DS, Theunissen B, Koolmees PA, Lenstra JA. On the history of cattle genetic resources. Diversity. 2014;6:705–50.
29. Yu Y, Nie L, He ZQ, Wen JK, Jian CS, Zhang YP. Mitochondrial DNA variation in cattle of south China: origin and introgression. Anim Genet. 1999;30(4):245–50.
30. Xia XT, Achilli A, Lenstra JA, Tong B, Ma Y, Huang YZ, Han JL, Sun ZY, Chen H, Lei CZ, et al. Mitochondrial genomes from modern and ancient Turano-Mongolian cattle reveal an ancient diversity of taurine maternal lineages in East Asia. Heredity (Edinb). 2021;126(6):1000–8.
31. Lai SJ, Liu YP, Liu YX, Li XW, Yao YG. Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation. Mol Phylogenet Evol. 2006;38(1):146–54.
32. Kalinowski ST. Do polymorphic loci require large sample sizes to estimate genetic distances? Heredity (Edinb). 2005;94(1):33–6.
33. Chen N, Xia X, Hanif Q, Zhang F, Dang R, Huang B, Lyu Y, Luo X, Zhang H, Yan H, et al. Global genetic diversity, introgression, and evolutionary adaptation of indicine cattle revealed by whole genome sequencing. Nat Commun. 2023;14(1):7803.
34. Wang MS, Zeng Y, Wang X, Nie WH, Wang JH, Su WT, Otecko NO, Xiong ZJ, Wang S, Qu KX et al. Draft genome of the gayal, Bos frontalis. GIGASCIENCE 2017;6(11):1–7.
35. Moran L, Mirault ME, Arrigo AP, Goldschmidt-Clermont M, Tissieres A. Heat shock of Drosophila melanogaster induces the synthesis of new messenger RNAs and proteins. Philos Trans R Soc Lond B Biol Sci. 1978;283(997):391–406.
36. Jakob U, Gaestel M, Engel K, Buchner J. Small heat shock proteins are molecular chaperones. J Biol Chem. 1993;268(3):1517–20.
37. Rogalla T, Ehrnsperger M, Preville X, Kotlyarov A, Lutsch G, Ducasse C, Paul C, Wieske M, Arrigo AP, Buchner J, et al. Regulation of Hsp27 oligomerization, chaperone function, and protective activity against oxidative stress/tumor necrosis factor alpha by phosphorylation. J Biol Chem. 1999;274(27):18947–56.
38. Pozo D, Vales-Gomez M, Mavaddat N, Williamson SC, Chisholm SE, Reyburn H. CD161 (human NKR-P1A) signaling in NK cells involves the activation of acid sphingomyelinase. J Immunol. 2006;176(4):2397–406.
39. Jaiswal S, Jamieson CH, Pang WW, Park CY, Chao MP, Majeti R, Traver D, van Rooijen N, Weissman IL. CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. Cell. 2009;138(2):271–85.
40. Willingham SB, Volkmer JP, Gentles AJ, Sahoo D, Dalerba P, Mitra SS, Wang J, Contreras-Trujillo H, Martin R, Cohen JD, et al. The CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for human solid tumors. Proc Natl Acad Sci U S A. 2012;109(17):6662–7.
41. Khan A, Khan MZ, Dou J, Xu H, Liu L, Zhu H, Wang Y. SOD1 gene silencing promotes apoptosis and suppresses proliferation of heat-stressed bovine granulosa cells via induction of oxidative stress. Vet Sci 2021;8(12).
42. Kambal S, Tijjani A, Ibrahim S, Ahmed MA, Mwacharo JM, Hanotte O. Candidate signatures of positive selection for environmental adaptation in indigenous African cattle: a review. Anim Genet. 2023;54(6):689–708.
43. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, et al. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 2009;324(5926):528–32.
44. Harringmeyer OS, Hoekstra HE. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. Nat Ecol Evol. 2022;6(12):1965–79.
45. Weissensteiner MH, Bunikis I, Catalan A, Francoijs KJ, Knief U, Heim W, Peona V, Pophaly SD, Sedlazeck FJ, Suh A, et al. Discovery and population genomics of structural variation in a songbird genus. Nat Commun. 2020;11(1):3403.
46. Verdugo MP, Mullin VE, Scheu A, Mattiangeli V, Daly KG, Maisano Delser P, Hare AJ, Burger J, Collins MJ, Kehati R, et al. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science. 2019;365(6449):173–76.
47. Lyu Y, Wang F, Cheng H, Han J, Dang R, Xia X, Wang H, Zhong J, Lenstra JA, Zhang H, et al. Recent selection and introgression facilitated high-altitude adaptation in cattle. Sci Bull (Beijing). 2024;69(21):3415–24.
48. 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):1297–1303.
49. Zhang Z, Wang A, Hu H, Wang L, Gong M, Yang Q, Liu A, Li R, Zhang H, Zhang Q, et al. The efficient phasing and imputation pipeline of low-coverage whole genome sequencing data using a high-quality and publicly available reference panel in cattle. Anim Res One Health. 2023;1(1):4–16.
50. Low WY. The case for bovine pangenome. Anim Res One Health. 2024;2(4):363–65.
51. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2(12):e190.

Publisher's note