RESEARCH



Admixture and selection offer insights for the conservation and breeding of Guyuan cattle

Shuang Liu^{1†}, Huixuan Yan^{2†}, Xue Feng^{1†}, Xiaoyu Luo², Yang Lv³, Chuzhao Lei^{2*}, Ningbo Chen^{2*} and Yun Ma^{1*}

Abstract

Background The admixture between taurine and indicine cattle increases breed diversity and provides new genetic resources for human and natural selection. The climate of northwestern China is typified by cold and arid conditions, and cattle in this region are primarily taurine breeds. However, Guyuan cattle inhabit a transitional zone in northwestern China, typified by semi-arid and semi-humid climates. It is hypothesized that Guyuan cattle have a little of indicine ancestry. These results suggest that Guyuan cattle are a valuable genetic resource.

Results We established a conservation population of Guyuan cattle in their native habitat. We found that Guyuan cattle were an admixture between 78.09% East Asian taurine (EAT) and 20.26% East Asian indicine (EAI) ancestries. The admixture in Guyuan cattle was dated to 255 years ago. Notably, we identified Guyuan cattle as a unique genetic resource, representing a transitional breed between northern and central Chinese cattle with distinct admixture proportions. We revealed that the selection signals in Guyuan cattle with excess EAT ancestry were associated with reproduction, immunity, body length, cold climate adaptation, pigmentation, muscle development, residual feed intake, and fat deposition and that the selection signals in Guyuan cattle with excess EAI ancestry were associated with disease resistance. Remarkably, we discovered valuable single nucleotide polymorphisms (SNPs) in the promoter regions of the *RBM39* and *NEK6* genes, which may play key roles in regulating muscle development and disease resistance.

Conclusions Our results suggest that Guyuan cattle are a newly identified genetic resource, and the native taurine and indicine ancestries in Guyuan cattle remain a valuable genetic resource of conservation and breeding.

Keywords Guyuan cattle, Conservation population, Admixture, Selection, Muscle development, Disease resistance

[†]Shuang Liu, Huixuan Yan and Xue Feng contributed equally to this work.

*Correspondence:
Chuzhao Lei
leichuzhao1118@nwafu.edu.cn
Ningbo Chen
ningbochen@nwafu.edu.cn
Yun Ma
mayun@nxu.edu.cn
¹ Key Laboratory of Molecular Cell Breeding of Ruminants in Ningxia Hui
Autonomous Region, College of Animal Science and Technology, Ningxia
University, Yinchuan 750021, China
² Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China
³ College of Animal Science and Technology, Jilin Agricultural University,



Changchun 130118, China

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Background

Cattle are one of the most important economic animals in the world, providing the high-quality protein-meat and milk that humans depend on for survival. According to the Food and Agriculture Organization of the United Nations (FAO), 1.4 billion cattle are kept worldwide. Despite their vast numbers, the expanding scope of human activities poses a significant threat to the genetic diversity of cattle, putting their breeds at risk of extinction. The FAO reports that cattle are an endangered species, with 17% of cattle breeds worldwide at risk and 254 breeds already extinct [1]. The widespread adoption of high-yield breeds, such as Holstein dairy cows and Angus beef cattle, has further exacerbated this situation. The loss of genetic diversity in cattle not only threatens the resilience and adaptability of breeds to environmental changes but also poses significant risks to global food security, especially in the context of climate change and emerging diseases. Genetic diversity serves as a critical buffer against these challenges, highlighting the urgency of developing effective conservation and breeding strategies based on genomic insights.

Based on genomic research, cattle worldwide are primarily divided into humpless taurine cattle and humped indicine cattle [2], which diverged more than 250,000 years ago [3]. Taurine cattle are predominantly found in the Northern Hemisphere; these cattle have adapted to temperate and cold climates and are known for their meat quality and growth characteristics [4]. In contrast, indicine cattle are distributed mainly in equatorial regions and the Southern Hemisphere; these cattle have adapted to tropical and subtropical climates and exhibit disease resistance and immune advantages [5]. Taurine×indicine cattle are widely bred in regions where temperate and tropical climates meet, exploiting advantage of breed complementarity and heterosis [6, 7]. Furthermore, genomic research consistently classifies cattle worldwide into five ancestral lineages: European taurine, Eurasian taurine, East Asian taurine, East Asian indicine, and South Asian indicine [8, 9].

Given the global situation of cattle genetic diversity, it is essential to explore the unique genetic resources within different regions. China, with its diverse cattle populations [10, 11], offers a valuable opportunity for research on the genetic diversity, ancestry, admixture and select of native cattle breeds. Native Chinese cattle have two types of ancestry: East Asian taurine (EAT) in the north, northwest, and Qinghai-Xizang of China; East Asian indicine (EAI) in the south of China. Different proportions of admixture between EAT and EAI cattle increase genetic diversity and provide novel genetic resources for human and natural selection. The taurine×indicine breeds that have been identified include Qinchuan [12], Zaosheng [13], Nanyang [15], Luxi [15], Jiaxian Red [15], Bohai Black [19], and Sanjiang [15] cattle.

Guyuan cattle are distributed in Guyuan City, Ningxia Hui Autonomous Region, in Northwest China (Fig. 1a). The geographic location of Guyuan cattle is unique. Guyuan cattle are located in the transitional zone between the northwest and north of China, as well as near the transitional zone between the north and south of China. As a result, Guyuan cattle inhabit a transitional zone between semi-arid and semi-humid climates. The highest rainfall occurs in August, averaging 109.73 ± 63.95 mm, with an average temperature of 19.35 ± 0.06 °C (Fig. 1b). The lowest temperature is recorded in January, averaging -7.52 ± 0.09 °C, with a minimal rainfall of 2.37 ± 3.36 mm (Fig. 1b). This geographic and climatic environment has contributed to the germplasm characteristics of Guyuan cattle. They are known for their climbing ability and disease resistance [19] and the admixture between taurine and indicine cattle in Guyuan cattle has occurred [23]. Previous studies on taurine \times indicine cattle mainly focused on certain well-known breeds, such as Qinchuan [12], Nanyang [15], Luxi [15], and Bohai Black [19] cattle, which are predominantly distributed in central China, particularly in the transitional zone between northern and southern China. These studies have mainly reported ancestry and admixture proportions, without the selection of different ancestries. Additionally, the selection of different ancestries in taurine×indicine cattle in the vast expanse of northwestern China has yet to be reported. Our study is the first to comprehensively analyze the genetic characteristics of Guyuan cattle from northwestern China, offering new insights into this regionally unique breed.

In this study, we established a conservation population of Guyuan cattle in their native habitat. By sequencing the genomes of 39 Guyuan cattle and collecting 165 genomes considering the geographical location and ancestry of Guyuan cattle (Fig. 1c). Through genomic analysis, we identified Guyuan cattle as a novel genetic resource, representing a transitional breed between northern and central Chinese cattle. Through ancestral inference and selective sweeps, we revealed that Guyuan cattle are a valuable genetic resource in reproduction, immunity, body length, cold climate adaptation, pigmentation, muscle development, residual feed intake, fat deposition, and disease resistance, offering valuable insights for their conservation and breeding strategies. Furthermore, we identified valuable single nucleotide polymorphisms (SNPs) in the missense mutations and promoter regions within the selection signals in Guyuan cattle. These results, exemplified by Guyuan cattle, provide valuable insights and resources for genetic improvement and breeding of taurine x indicine cattle.



Fig. 1 Geographic distribution of Guyuan cattle and other breeds used in this study. Geographical distribution of the 21 breeds, comprising 204 individuals. **a** Guyuan cattle and their natural habitat. **b** Monthly average rainfall and temperature in the geographic regions of Guyuan cattle and its neighboring breeds (Zaosheng, Qinchuan, and Mongolian cattle). **c** Geographical distribution of 21 breeds, comprising 204 individuals. *AGS*, Angus; *HER*, Hereford; *SIM*, Simmental; *GEL*, Gelbvieh; *HAN*, Hanwoo; *YB*, Yanbian; *XZT*, Xizang; *CDM*, Qaidam; *AX*; Anxi; *MG*, Mongolian; *GY*, Guyuan; *ZS*, Zaosheng; *QC*, Qinchuan; *XN*, Xiangnan; *WZ*, Weizhou; *HN*, Hainan; *RSD*, Red Sindhi; *DN*, Dhanni; *CLT*, Cholistani

Results

Population genetic structure and genomic diversity

The population genetic structure analyses, including neighbor-joining (NJ) tree, principal component analysis (PCA), and ADMIXTURE, consistently demonstrated that Guyuan cattle represent a distinct genetic resource, clearly separated from geographically adjacent breeds such as Mongolian, Zaosheng, and Qinchuan cattle. Their history dates back to the Northern Wei Dynasty (~1,500 years ago) (Additional file 1: Fig. S1). The NJ tree revealed that Guyuan cattle forms a unique branch, positioned closer to EAT and EAI cattle but maintaining a distinct structure, highlighting its independent genetic lineage (Fig. 2a). The PCA further supported this finding which Guyuan cattle are the admixture between EAT and EAI cattle, but distinct from Zaosheng and Qinchuan cattle, reflecting its unique genetic composition (Fig. 2b). The ADMIXTURE analysis identified that Guyuan cattle have 20.26% EAI ancestries, but Zaosheng and Qinchuan cattle have 25.43% and 25.97% EAI ancestries, displaying its unique ancestral components (Fig. 2c). These results not only demonstrated the unique genetic composition of Guyuan cattle but also suggested that their genetic structure was influenced by a combination of EAT and EAI ancestries, highlighting their distinct evolutionary history and genetic differentiation from neighboring breeds.

The nucleotide diversity, number of total SNPs, and number of breed-specific SNPs of Guyuan cattle were intermediate between the values of EAT and EAI cattle (Fig. 3a, b). The maximum linkage disequilibrium (LD) coefficient (r^2) of Guyuan cattle was 0.5403, which was lower than those of EAT (0.6569) cattle but higher than that of EAI (0.4426) cattle (Fig. 3c). Short runs of homozygosity (ROHs) (< 0.5 Mb) accounted for 94.48% in Guyuan cattle (Fig. 3d), compared to 95.22% in EAT cattle and 97.17% in EAI cattle. This findings also suggested that the genomic diversity of Guyuan cattle was intermediate between that of taurine and indicine cattle. Consistently, these results also suggested that the genomic diversity of Guyuan cattle was intermediate between that of EAT and EAI cattle, further supporting the notion of Guyuan cattle as a unique admixture.

Collectively, these results provided strong evidence that Guyuan cattle are a unique genetic resource, typified by significant differentiation from neighboring breeds as well as distinct genetic structure and genomic diversity. Furthermore, their genetic structure, influenced by the combination of EAT and EAI ancestries, may contribute to their climbing ability and disease resistance,



Fig. 2 Population genetic structure of Guyuan cattle. **a** NJ tree of all individuals. **b** PCA showed PC1 and PC2 for all individuals. **c** Admixture model results for *K*=2 and *K*=4. *AGS*, Angus; *HER*, Hereford; *SIM*, Simmental; *GEL*, Gelbvieh; *HAN*, Hanwoo; *YB*, Yanbian; *XZT*, Xizang; *CDM*, Qaidam; *AX*; Anxi; *MG*, Mongolian; *GY*, Guyuan; *ZS*, Zaosheng; *QC*, Qinchuan; *XN*, Xiangnan; *WZ*, Weizhou; *HN*, Hainan; *RSD*, Red Sindhi; *DN*, Dhanni; *CLT*, Cholistani

highlighting their potential for adaptation to local environments. These findings underscore the importance of Guyuan cattle for conservation and further genetic studies.

Ancestry inference and missense mutations

The proportion of EAI ancestry in Guyuan, Zaosheng, and Qinchuan cattle, previously estimated by ADMIX-TURE, was further estimated using RFMix and Admixtools (Table 1). These results consistently showed that the proportion of EAI ancestry decreased geographically from south to north among these cattle breeds. Ancestry segments of EAT and EAI in Guyuan cattle were inferred from LOTER. Ultimately, 8,535 EAT segments and 2,652 EAI segments were retained (Fig. 4a and Additional file 2: Table S1). The NJ tree of these segments confirmed the results (Fig. 4b, c). For the excess EAT ancestry segments in Guyuan cattle, 1,245 genes were annotated, which were enriched in pathways related to meat quality, growth and development (Fig. 4d). For the excess EAI ancestry segments in Guyuan cattle, 357 genes were annotated, which were enriched in pathways related to disease resistance (Fig. 4e). The admixture in Guyuan cattle occurred around 51 generations ago, based on LD decay patterns analyzed with ALDER. Assuming a generation time of 5 years for cattle [15], this corresponds to approximately 255 years ago.

A total of 102 missense mutations in 74 genes were identified with a frequency of missense mutations greater than 0.8 or less than 0.1 in Guyuan and EAT cattle and less than 0.1 or greater than 0.8 in EAI cattle (Additional file 2: Table S2). A total of 46 missense mutations in 41 genes were identified with a frequency of missense mutations greater than 0.45 or less than 0.1 in Guyuan and EAI cattle and less than 0.1 or greater than 0.35 in EAT cattle (Additional file 2: Table S2).

Selection signals in Guyuan cattle with excess EAT ancestry

We first explored continuous sweeps of selection regions using integrated haplotype scores (*iHS*). The top 2% of regions containing a proportion of SNPs with $||iHS| \ge 2$ were identified, yielding a total of 2,484 potential selection regions (Additional file 1: Fig. S2). We then calculated the fixation index ($F_{\rm ST}$) and cross population extended haplotype homozygosity (XP-EHH) values between Guyuan and EAI cattle. After the *Z* tests were conducted, 8,166 and 5,185 windows, respectively, were retained as potential selection regions (Additional file 1: Fig. S3a, b). A total of 105 windows identified by *iHS*, $F_{\rm ST}$, and XP-EHH overlapped. We further considered



Fig. 3 Genomic diversity of Guyuan cattle. **a** The genome-wide distribution of nucleotide diversity in each breed is presented in 50 kb windows with a 20 kb steps. **b** Number of SNPs identified in each breed with respect to the reference genome. High and low bars represent the numbers of all SNPs and breed-specific SNPs, respectively. **c** Genome-wide average LD decay estimated for each breed. **d** ROHs patterns of all individuals from each cattle geographic groups. *EUT*, European taurine; *EAT*, East Asian taurine; *GY*, Guyuan; *ZS*, Zaosheng; *QC*, Qinchuan; *EAI*, East Asian indicine; *SAI*; South Asia indicine

Table 1 The proportion of EAI ancestry in Guyuan, Zaosheng,
and Qinchuan cattle was estimated via ADMIXTURE, RFMix, and
Admixtools

Cattle breed	ADMIXTURE (%)	RFMix (%)	Admixtools (%)		
Guyuan	20.26	25.94	21		
Zaosheng	25.43	27.70	25		
Qinchuan	25.97	29.85	26		

the proportion of EAT ancestry (Additional file 1: Fig. S3c) across the entire genome (EAT ancestry > 88.95%) and combined adjacent windows, ultimately retaining 16 genomic regions containing 21 genes (Table 2).

The selection signals in Guyuan cattle with excess EAT ancestry include *CAPN7* [19] and *GALNTL6* [15, 23], associated with reproduction; *SLAMF1*, *CD84*, and *SLAMF6*, associated with immunity [19]; *CACNA1C*, associated with body length [23]; *SPAS7*, associated with cold climate adaptation; *ASIP* and *AHCY*, associated with pigmentation [9]; *CPNE1* [15, 19], *NFS1* [23], *ROMO1*





Identification of the local segments in which proportions of a certain ancestry were significantly higher than the proportion in the whole genome in Guyuan cattle. **a** Distribution of the local segments with proportions of EAT and EAI ancestries. **b** NJ tree of Guyuan cattle with excessive EAT ancestry, EAT, and EAI cattle. **c** NJ tree of Guyuan cattle with excessive EAI ancestry, EAT, and EAI cattle. **d** The KEGG pathways from the enrichment analysis of genes with excessive EAT ancestry. **e** The KEGG pathways from the enrichment analysis of genes with excessive EAI ancestry. **f** The significant difference in the frequency of missense mutations between the group (Guyuan and EAI cattle) and the group (EAT cattle)

BTA	Regions (Mb)	No. of windows	Proportion of SNPs with <i>iHS</i> ≥2 (%)	F _{ST}	XP-EHH	EAT Ancestry (%)	Genes identified	Previous studies
1	152.18–152.25	2	71.00%	0.54	1.59	89.51%	LOC112448102 CAPN7	Reproduction (Muhammad et al. 2018)
3	8.96-9.09	5	97.11%	0.53	5.22	92.09%	SLAMF1 CD84	Immune (Qi. 2021)
3	9.16-9.21	1	86.25%	0.52	4.70	90.08%	SLAMF6	Immune (Qi. 2021)
5	108.48–108.53	1	44.34%	0.52	1.57	91.24%	CACNA1C	Body length (Chen et al. 2020)
6	70.00-70.05	1	41.11%	0.50	2.42	89.31%	_	—
8	4.26-4.31	1	35.05%	0.53	1.91	92.34%	GALNTL6	Reproduction (Borowska et al. 2018; Ghoreishifar et al. 2023)
11	14.70–14.75	1	91.75%	0.57	1.59	90.43%	SPAST	Cold climate adaptation (Wang et al. 2022; Mahar et al. 2024)
13	15.58–15.63	1	45.36%	0.54	1.70	90.16%	_	—
13	63.54–63.59	1	35.34%	0.71	2.03	91.12%	_	—
13	63.58–63.75	7	61.59%	0.70	3.02	92.17%	ASIP AHCY	Pigmentation (Chen et al. 2023)
13	64.80–65.03	10	54.22%	0.60	2.65	94.55%	SPAG4 LOC526745 CPNE1 R BM12 NFS1 ROM01 RBM3 9 PHF20	Muscle development (Her- nandez et al. 2019; Saha et al. 2014; Anderson et al. 2016; Jones et al. 2023; Lee et al. 2020; Chen et al. 2022)
14	21.86–21.91	1	62.69%	0.51	2.31	91.02%	ATP6V1H	Residual feed intake (de Las et al. 2019; Uemoto et al. 2020; Brunes et al. 2021)
14	22.06-22.19	5	63.40%	0.56	1.91	92.00%	—	_
14	22.82-22.93	4	47.58%	0.59	2.45	91.59%	XKR4	Residual feed intake (Lind- holm-Perry et al. 2012)
16	23.46-23.51	1	51.81%	0.51	2.44	94.24%	EPRS	Fat deposition (Arif et al. 2017; Hay et al. 2018)
29	46.74-46.79	1	63.74%	0.53	1.84	89.55%	—	_

Table 2	Candidate	regions	with an	excess of	[:] EAT	ancestry	in Guy	/uan (cattle.	Note:	Bos ta	aurus	autoso	me (BTA)
		9												· · ·

[15], *RBM39* [19], and *PHF20* [23], associated with muscle development; *ATP6V1H* [15, 19, 23] and *XKR4* [15], associated with residual feed intake; and *EPRS*, associated with fat deposition [19, 23].

The four strongest selection signals in Guyuan cattle with excess EAT ancestry (Additional file 1: Fig. S4) were *Bos taurus* autosome 3 (BTA3) (8.96–9.09 Mb), BTA3 (9.16–9.21 Mb), BTA13 (63.58–63.75 Mb), and BTA13 (64.80–65.03 Mb). The average proportions of EAT ancestry in these regions were 90.08%, 92.09%, 92.17%, and 94.55%, respectively. These regions in Guyuan cattle had high proportions of SNPs with $|iHS| \ge 2$, high proportions of EAT ancestry, high F_{ST} values, and low Tajimas' *D* values (Fig. 5a and Additional file 1: Fig. S5a). Haplotype sharing was observed between Guyuan and EAT cattle in these regions (Fig. 5b and Additional file 1: Fig. S5b). The core variants of these regions were rs137179098 (BTA3: 9,010,601), rs133600619 (BTA3: 9,170,093), rs110917065 (BTA13: 63,701,514), and rs449336579 (BTA13: 65,039,842) in the extended

⁽See figure on next page.)

Candidate regions of BTA13 (63.58–63.75 Mb) and BTA3 (64.80–65.03 Mb) in Guyuan cattle with an excess of EAT ancestry. **a** For the candidate regions, the proportions of SNPs with $|iHS| \ge 2$ and EAT ancestry in Guyuan cattle were calculated in 5 kb windows with a 2 kb step, along with the F_{ST} values between Guyuan and Hanwoo or EAI cattle and the Tajimas' *D* values in Guyuan and EAI cattle. **b** Haplotypes in the candidate regions (major alleles shown in orange, minor alleles in blue). **c** Core SNPs of the candidate regions in the EHH test. **d** SNPs in the promoter of the *RBM39* gene identified by ATAC_seq and H3K4me3 ChIP_seq in bovine muscle tissues. **e** Expression of the *RBM39* gene identified by RNA-seq in bovine different muscle tissues



(See legend on previous page.)

haplotype homozygosity (EHH) test (Fig. 5c and Additional file 1: Fig. S5c).

Two missense mutations within the selection signals in Guyuan cattle with excess EAT ancestry, rs449336579 (PHF20; BTA13: T65,039,842C; Val459Ala) and rs136018834 (ATP6V1H; BTA14: G21,874,515A; Val-279Ile), were significantly different between taurine cattle and indicine cattle. The distributions of the tow SNPs in taurine, indicine, and taurine×indicine cattle worldwide reveaed same results (Additional file 2: Table S4). We downloaded the structure of bovine ATP6V1H and predicted the structures of both the wild-type and mutant proteins using AlphaFold2. This mutation altered the structure of ATP6V1H (Additional file 1: Fig. S6).

Guyuan cattle are known for climbing ability. All genes in BTA13 (64.80–65.03 Mb) are associated with muscle development. To identify potentially functional SNPs in promoter regions, we focused on SNPs whose $F_{\rm ST}$ values between Guyuan and EAI cattle exceeded the genomewide level (P<0.05). Furthermore, we overlapped these SNPs with the regions identified by ATAC_seq and H3K4Me3 ChIP_seq in bovine muscle tissues. For BTA13 (64.80–65.03 Mb), we found three potentially functional SNPs in the promoter region of the *RBM39* gene in bovine muscle tissues (Fig. 5d). RNA-seq revealed that the *RBM39* gene was expressed in bovine different muscle tissues (Fig. 5e).

Selection signals in Guyuan cattle with excess EAI ancestry

We calculated the $F_{\rm ST}$ and XP-EHH values between Guyuan and Hanwoo cattle. After conducting *Z* tests, we identified 8,276 and 5,516 windows, respectively, as potential selection regions (Additional file 1: Fig. S7a, b). Among these, 538 windows overlapped across the three methods (*iHS*, $F_{\rm ST}$, and XP-EHH). By further considering the proportion of EAI ancestry (Additional file 1: Fig. S7c) across the genome (EAI ancestry > 45.69%) and merging adjacent windows, we retained three genomic regions (Table 3).

BTA11 (95.20–95.37 Mb) is the strongest selection signal in Guyuan cattle with excess EAI ancestry (Additional file 1: Fig. S8), including part of the *NEK6* gene, which has been associated with disease resistance [15]. The average proportion of EAI ancestry in the region is

65.67%. The region had high proportion of SNPs with $|iHS| \ge 2$, high proportion of EAI ancestry, high F_{ST} values, low $\theta \pi$ values, and low Tajimas' *D* values (Fig. 6a). Haplotype sharing was observed between Guyuan and EAI cattle in the region (Fig. 6b). The core variant of the region was rs208403297 (BTA11: 95,322,218) in the EHH test (Fig. 6c).

Guyuan cattle are known for disease resistance. The spleen serves as a reservoir of immune cells. For BTA11 (95.20–95.37 Mb), we overlapped SNPs whose F_{ST} values between Guyuan and Hanwoo cattle that exceeded the genome-wide level (P<0.05) with the regions identified by ATAC_seq and H3K4Me3 ChIP_seq in bovine spleen tissues. We found a potentially functional SNP in the promoter region of the *NEK6* gene (Fig. 6d). RNA-seq demonstrated robust expression of the *NEK6* gene in tissues highly enriched with immune cells (Fig. 6e)

Discussion

Most reported taurine×indicine cattle breeds result from the admixture between European taurine and South Asian indicine cattle [19]. In contrast, China's native taurine×indicine cattle breeds are the admixture between EAT and EAI cattle [8]. Approximately 4,000-5,000 years ago, taurine cattle migrated eastward from their original domestication center in the Near East to northern China. Around 3,000 years ago, indicine cattle from India migrated to China, leading to the emergence of admixture. The geographical regions of China are divided into the northern, southern, northwestern, and Qinghai-Xizang regions. Cattle breeds in northern, northwestern and Qinghai-Xizang China are mostly taurine cattle, such as Yanbian [8], Mongolian [23], Qaidam [15], Anxi [19], and Xizang [48] cattle. In contrast, the breeds in southern China are mostly indicine cattle, such as Hainan [23] cattle. The breeds in the transitional zone between northern and southern China are often taurine×indicine cattle, such as Qinchuan [12], Nanyang [15], Luxi [15], and Bohai Black [19] cattle. Compared with taurine cattle breeds from northwestern China, such as Mongolian [23] and Anxi [19] and cattle, Guyuan cattle have the highest proportion of EAI ancestry. In contrast, compared with taurine×indicine cattle breeds from the central plains of China, such as Qinchuan [12], Nanyang [15], Luxi

Table 3 Candidate regions with an excess of EAI ancestry in Guyuan cattle. Note: Bos taurus autosome (BTA)

BTA	Regions (Mb)	No. of windows	Proportion of SNPs with <i>iHS</i> ≥2 (%)	F _{ST}	XP-EHH	EAI ancestry (%)	Genes identified	Previous studies
6	51.78-51.83	1	38.60%	0.25	2.47	49.85%	_	
11	95.20–95.37	7	52.79%	0.22	2.95	65.67%	NEK6	Disease resistance (Kumar et al. 2022)
20	28.92-28.99	2	57.50%	0.17	1.57	57.16%	—	



Fig. 6 Candidate region of BTA11 (95.20–95.37 Mb) in Guyuan cattle with an excess of EAI ancestry. **a** For the candidate region, the proportions of SNPs with $|iHS| \ge 2$ and EAI ancestry in Guyuan cattle were calculated in 5 kb windows with a 2 kb step, along with the F_{ST} values between Guyuan and Hanwoo or EAI cattle and the Tajimas' *D* values in Guyuan and Hanwoo cattle. **b** Haplotypes in the candidate regions (major alleles shown in orange, minor alleles in blue). **c** Core SNP of the candidate region in the EHH test. **d** A SNP in the promoter of the *NEK6* gene identified by ATAC_ seq and H3K4me3 ChIP_seq in bovine spleen tissues. **e** Expression of the *NEK6* gene identified by RNA-seq in in bovine tissues highly enriched with immune cells

[15], and Bohai Black [19], Guyuan cattle have the lowest proportion of EAI ancestry. Guyuan cattle inhabit a transitional zone in northwestern China, which may have contributed to their unique proportion of EAI ancestry, potentially leading to differences in traits such as cold climate adaptation and disease resistance.

The admixture between taurine and indicine cattle has played a pivotal role in shaping the genetic diversity of cattle breeds, while simultaneously providing novel genetic resources for both human and natural selection. Taurine×indicine cattle, commercially referred to as composite cattle, have become integral to the global beef industry due to their breed complementarity and heterosis. However, such admixture is not without cost, as genomic incompatibilities can arise, potentially compromising reproductive fitness [7]. Over time, the forces of human and natural selection have acted to mitigate these challenges by favoring the retention of genomic regions enriched with advantageous variations from specific ancestries, while purging regions carrying deleterious variants. This select has not only alleviated the detrimental effects of admixture but also optimized the genetic architecture and evolutionary potential of these hybrid populations.

Selection played a role in shaping the proportions of taurine and indicine ancestries in Guyuan cattle. The admixture between indicine and taurine cattle is a key feature in the breed formation of Guyuan cattle. The ancestry in Guyuan cattle was found to be more skewed toward taurine. SLAMF1, CD84 (SLAMF5), SLAMF6, ASIP, AHCY, SPAG4, CPNE1, NFS1, RBM12, ROMO1, RBM39, and PHF20 were identified as strong selected genes with excess EAT ancestry in Guyuan cattle. SLAMF1, CD84 (SLAMF5), and SLAMF6 in BTA3 (8.96-9.09 Mb) and BTA3 (9.16–9.21 Mb) are well-known immune response genes that play crucial roles in resistance to Mycoplasma infections in cattle [15, 19]. It has been demonstrated that SLAMF1, CD84 (SLAMF5), and SLAMF6 synergistically regulate humoral immune responses through enhanced antibody responses to T-dependent and T-independent antigens in mice [48]. This mechanism may contribute to the rapid pathogen response in Guyuan cattle. ASIP and AHCY in BTA13 (63.58-63.75 Mb) are recognized pigmentation genes [9]; the red-brown coloration of Guyuan cattle may be associated with SNPs at the ASIP locus that cause a darker coat color. SPAG4, CPNE1, NFS1, RBM12, ROMO1, RBM39, and PHF20 in BTA13 (64.80-65.03 Mb) were identified as strong selected genes with excess EAT ancestry in Guyuan cattle. Notably, these genes also showed a strong association with primal cut lean traits in Canadian beef cattle (taurine cattle breeds), as revealed by a genome-wide association study (GWAS) [23]. Moreover, functional studies highlight their crucial

roles in muscle development. CPNE1 and RBM12 share a common promoter, and CPNE1 regulates myogenesis through the PERK-eIF2α pathway mediated by endoplasmic reticulum stress [15, 19]. NFS1 acts as a sulfur donor interacting with the J-protein HSCB, and its dysfunction impairs mitochondrial electron transport chain function, leading to muscle atrophy [23]. The mitochondrial gene *ROMO1* is involved in the reprogramming of muscle cells [15]. RBM39 interacts with ZFP106 to regulate myogenesis, and its deficiency results in muscle atrophy [19]. PHF20 binds to the YY1 promoter which is essential for myogenic differentiation in vitro and in vivo [23]. Therefore, we infer that SNPs in BTA13 (64.80-65.03 Mb) are associated with muscle development and may contribute to the strong climbing ability in Guyuan cattle. disease resistance is a characteristic of indicine cattle. NEK6 in BTA11 (95.20-95.37 Mb) was identified as a strong selected gene with excess EAI ancestry in Guyuan cattle. NEK6 was associated with disease resistance, including resistance to African trypanosomiasis (Nagana) [15], Marek's disease [15], ciliary dyskinesia [19], and tuberculosis [48]. Previous study has shown that NEK6 can modulate cell cycle progression and apoptosis in immune cells when infected by pathogens [48], which could enhance the immune response against pathogens in Guyuan cattle. Therefore, SNPs in BTA11 (95.20-95.37 Mb) may offer advantageous in disease resistance traits.

On the one hand, missense mutations, by altering the amino acid sequence of proteins, can enhance or impair protein functionality, thereby playing a pivotal role in the development of biological traits and adaptive evolution. The missense mutation, rs136018834 (ATP6V1H; BTA14: G21,874,515A; Val279Ile), is within the selection signals in Guyuan cattle with excess taurine ancestry. The frequency of rs136018834 is significantly different between taurine cattle and indicine cattle. Additionally, the spatial structure of ATP6V1H protein from bovine brain has been characterized [23]. Using Alpha-Fold2, we further predicted that this mutation might alter the spatial structure of ATP6V1H. Divergent selection experiments have demonstrated that ATP6V1H is associated with feed conversion efficiency in mice [15]. Genome-wide association studies have also confirmed that ATP6V1H is associated with residual feed intake in Angus [15], Japanese Black [19], and Nellore [23] cattle. Additionally, ATP6V1H is regarded as an oxidative phosphorylation gene, regulating cellular energy metabolism and being closely associated with insulin secretion and glucose metabolism [19]. The missense mutation, rs136018834 (ATP6V1H; BTA14: G21,874,515A; Val-279Ile) may potentially affect its function and plays a critical role in optimizing feed efficiency traits in Guyuan cattle and other taurine cattle breeds.

On the other hand, SNPs in promoter regions finely regulate gene expression patterns, influencing the spatiotemporal specificity of gene function. This provides the genetic foundation for shaping the genetic diversity and complex traits of species. Cis-regulatory elements, especially promoters, of mammalian complex traits exhibit strong conservation across species and breeds [23, 48]. For the conservation and breeding of Guyuan cattle, this study integrated ATAC_seq, H3K4Me3 ChIP seq, and RNA-seq data from other cattle breeds to identify functional SNPs in the promoter regions of the RBM39 and NEK6 genes. Given the functional mechanisms of RBM39 and NEK6 genes [19, 48], these SNPs in the promoter regions may be associated with muscle development and disease resistance in Guyuan cattle, respectively.

The identification of these SNPs in missense mutations and promoter regions offers insights into the genetic basis of complex traits and provides critical theoretical support for the conservation and breeding of Guyuan cattle as a new genetic resource.

Conclusions

Guyuan cattle were identified as a new genetic resource and the admixture between EAT and EAI ancestries. They represent an important genetic resource for understanding the cattle breeds inhabit a transitional zone between semi-arid and semi-humid climate. The selected genes with excess EAT ancestry in Guyuan cattle were associated with immunity, pigmentation, and muscle development. The selected genes with excess EAI ancestry were associated with disease resistance. Our results suggest that the native taurine and indicine ancestries in Guyuan cattle offer valuable insights for conservation and breeding for their conservation and breeding.

Methods

Ethics statement

The study was approved by the Institutional Animal Care and Use Committee of Ningxia University (NXU-A-2023-162) and adhered to the recommendations of the Regulations for the Administration of Affairs Concerning Experimental Animals of China. Specific consent procedures were not required for this study according to the recommendation of the Regulations for the Administration of Affairs Concerning Experimental Animals of China. All operations and experimental procedures complied with the National Standard of Laboratory Animal Guidelines for Ethical Review of Animal Welfare (GB/T 35892–2018) and the Guide for the Care and Use of Laboratory Animals: Eighth Edition.

Samples and sequencing

We collected blood samples from 39 Guyuan cattle at the breeding farm, established by Guyuan Fumin Agricultural Technology Co., Ltd. Genomic DNA was extracted by the standard phenol–chloroform method and then sent to Shenzhen BGI Genomics Co., Ltd., where wholegenome resequencing was performed on the DNB-SEQ-T7 instrument, utilizing the DNBSEQ platform. The sequencing was conducted in PE150 mode (150 bp paired-end reads), with an average sequencing depth of $15 \times$. In accordance with the methodology described by Chen et al. [8], we used 165 samples from 20 breeds as a control group.

Identification of SNPs

The raw reads were filtered for the clean reads using Fastp (v0.20.0) [15] with the parameters "-n 10 -q 20 -u 40". The clean reads were then trimmed using Trimmomatic (v0.38) [19] with the parameters "LEADING:20 TRAIL-ING:20 SLIDINGWINDOW:3:15 AVGQUAL:20 MIN-LEN:35 TOPHRED33". The trimmed reads were were aligned to the cattle reference genome (ARS-UCD1.2_ Btau5.0.1Y) using BWA-MEM (v0.7.13) [48] with default parameters. Duplicates reads were removed using the SortSam and MarkDuplicates tools in Picard (v2.20.2) (https://broadinstitute.github.io/picard/). Alignment rates and average depths for all samples were calculated using Qualimap (v2.2) [23]. SNPs for all samples were identified using the HaplotypeCaller, GenotypeGVCFs, and SelectVariants tools in GATK (v.3.8) [15]. SNPs were filtered using the VariantFiltration tools in GATK (v.3.8) with the parameters: (1) mean sequencing depth for all samples > $1/3 \times and < 3 \times$; (2) QD < 2; (3) FS > 60.0; (4) MQ < 40; (5) MQRankSum < -12.5; (6) ReadPosRank-Sum < -8; (7) SOR > 3.0. Biallelic SNPs for all samples were identified and merged using the SelectVariants and CatVariants tools in GATK4 (v4.0) [15].

Phylogenetic and population structure

The pairwise genetic distances between SNPs were calculated using PLINK (v1.90) [19], and the distance matrix was then used in MEGA (v11.0) [48] to construct the neighbor-joining (NJ) tree. SNP pruning was conducted using PLINK (v1.90) with the parameters "-maf 0.05" to exclude SNPs with a minor allele frequency below 5%, and "-indep-pairwise 50 10 0.1" to pruning SNPs with high linkage disequilibrium ($r^2 > 0.1$) within a window size of 50 SNPs and a step size of 10 SNPs [8, 9]. Principal component analysis (PCA) and population structure analysis were conducted using the pruned SNPs, with PCA performed using the smartPCA package in EIGEN-SOFT (v5.0) [23] and population structure analysis performed using ADMIXTURE (v1.3.0) [15].

Genomic diversity

The nucleotide diversity ($\theta\pi$) for each breed was estimated using VCFtools (v0.1.17) [19], with a window size of 50 kb and a step size of 20 kb. Breed-specific SNPs were identified using PLINK. Linkage disequilibrium (LD) between SNPs was calculated using PopLDdecay [48]. Runs of homozygosity (ROHs) were identified using the "-homozyg" option in PLINK, with the following parameters: (1) homozyg-density 200; (2) homozyg-snp 50; (3) homozyg-window-snp 100; (4) homozyg-kb 100; (5) homozyg-gap 1000; (6) homozyg-window-threshold 0.05; (7) homozyg-window-het 1.

Ancestry inference

Phase and imputation of the raw genotype data were performed using Beagle (v4.1) [23]. The ancestry proportions of Guyuan cattle were further validated using Admixtools (v8.0.1) [15] and RFmix (v2.03) [19]. The time of admixture in Guyuan cattle was estimated using ALDER [48], with a minimum genetic distance (mindis) of 0.5 cM. On the basis of the population genetic structure, EAT (East Asian taurine) and EAI (East Asian indicine) were selected as reference groups. Lengths and proportions of ancestry segments were inferred using LOTER [23]. KEGG enrichment analysis of genes was conducted using KOBAS (http://kobas.cbi.pku.edu.cn/).

Detection of missense mutations and analysis of protein structure

SNPs were annotated using SNPeff [15] to identify those associated with missense mutations. The distributions of these SNPs in taurine, indicine, and taurine×indicine cattle worldwide were retrieved from the BDGV database (http://animal.omics.pro/code/index.php/BosVar) [19]. The protein 3D structure files were downloaded from the PDB database (https://www.rcsb.org/). The 3D structure of the protein sequence was predicted using AlphaFold2 [48]. PyMOL was employed to compare the spatial structures of the mutant and wild-type proteins, with a focus on conformational changes in local structures.

Detection of the selection signals

The raw genotype data were merged with the swamp buffalo genotype data, followed by phasing and imputation for *iHS*. The genotype was polarized using VCFbyOutgroup.py (https://github.com/kullrich/bio-scrip ts), with the swamp buffalo designated as the outgroup to as certain ancestral alleles [9, 23]. iHS was calculated using Selscan (v2.0) [85] with the parameter "-f 0.01" and the proportion of SNPs with $|iHS| \ge 2$ was determined in 50 kb windows with a 20 kb step size. The fixation index (F_{ST}) was calculated in 50 kb windows with a 20 kb step size using VCFtools, and XP-EHH was similarly calculated using Selscan. Z tests were conducted for each window. Phased genotype data with a minor allele frequency below 0.2 were filtered out, and haplotypes of candidate regions were visualized using Python's matplotlib and seaborn libraries. For phased genotype data containing ancestral and derived alleles, EHH analysis was performed on the candidate regions using Rehh [15] to identify core variants and plot the EHH curves.

Analysis of ATAC_seq, H3 K4Me3 ChIP-seq, and RNA-seq data

The ATAC_seq and H3K4Me3 ChIP_seq data were processed using the following pipeline: the clean reads were aligned to the ARS-UCD1.2_Btau5.0.1Y using Bowtie2 (v2.4.5) [19]; adapter sequences were removed with Trim Galore (v0.6.10) (https://github.com/FelixKrueger/TrimGalore); the bam files were sorted using SAMtools (v1.17) [48], and duplicates were removed with Picard (v2.02) (https://broadinstitute.github.io/picard); peaks were called for each replicate using MACS2 (v2.2.7.1) [85]; and biological replicates were merged using IDR (v2.0.4.2) [85]. These ATAC_seq and H3K4Me3 ChIP_seq data were from the NCBI with the Bioproject accession number PRJNA665194 [23]. RNA-seq data were retrieved from the RGD2.0 database (http://animal.omics.pro/code/index.php/RGD) [85].

Abbreviations

BTA	Bos taurus autosome
EAT	East Asian taurine
EAI	East Asian indicine
SNPs	Single nucleotide polymorphisms
iHS	Integrated haplotype score
F _{ST}	Fixation index
XP-EHH	Cross population extended haplotype homozygosity
EHH	Extended haplotype homozygosity

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12915-025-02213-y.

Additional file 1: Fig. S1. The Guyuan painted pottery cattle during the Northern Wei Dynasty (~1,500 years ago). Fig. S2. Distribution of SNP proportions with $|iHS| \ge 2$ in Guyuan cattle, analyzed across each 50-kb window with a 20-kb step. Dashed line indicates the threshold of the highest 2% of SNP proportions with $|iHS| \ge 2$. Fig. S3. Distribution of the F_{ST} and XP-EHH values between Guyuan and EAI cattle, and EAT ancestry proportions in Guyuan cattle, analyzed across each 50-kb window with a 20-kb step. Fig. S4. The selection signals in Guyuan cattle with an excess of EAT ancestry. The proportion of SNPs with $|iHS| \ge 2$ in Guyuan cattle; the F_{ST} and XP-EHH values between Guyuan and EAI cattle. Points with EAT ancestry significantly higher than the genome-wide levelare marked with larger dots. Dashed lines indicate the thresholds of the highest 2% of SNP proportions with $|iHS| \ge 2$ in Guyuan cattle. Fig. S4. The selection signals are the thresholds of the highest 2% of SNP proportions with $|iHS| \ge 2$ in Guyuan cattle and the F_{ST} values higher than the genome-wide levels for the marked with larger dots. Dashed lines indicate the thresholds of the highest 2% of SNP proportions with $|iHS| \ge 2$ in Guyuan cattle. Fig. S5.

The candidate regions of BTA3 (8.96-9.09 Mb) and BTA3 (9.16-9.21 Mb) in Guyuan cattle with an excess of EAT ancestry. Fig. S6. The comparison of protein structures. The blue structure represents the crystal structure of the bovine ATP6V1H protein, the yellow structure represents the 3D structure of the wild-type ATP6V1H protein predicted by AlphaFold2, and the red structure represents the 3D structure of the mutant ATP6V1H protein. Fig. S7. Distribution of the F_{ST} and XP-EHH values between Guyuan and Hanwoo cattle, and EAI ancestry proportions in Guyuan cattle, analyzed across each 50-kb window with a 20-kb step. Fig. S8. The selection signals in Guyuan cattle with an excess of EAI ancestry. The proportion of SNPs with $|iHS| \ge 2$ in Guyuan cattle; the F_{ST} and XP-EHH values between Guyuan and Hanwoo cattle. Points with EAI ancestry significantly higher than the genome-wide levelare marked with larger dots. Dashed lines indicate the thresholds of the highest 2% of SNP proportions with $|iHS| \ge$ 2 in Guyuan cattle and the F_{ST} values higher than the genome-wide level between Guyuan and Hanwoo cattle.

Additional file 2: Table S1. The genes with excess taurine or indicine ancestry segments detected by LOTER. Table S2. The significant difference in the frequency of missense mutations in Guyuan cattle between the group (Guyuan and East Asian taurine cattle) and the group (East Asian indicine cattle). Table S3. The significant difference in the frequency of missense mutations in Guyuan cattle between the group (Guyuan and East Asian indicine cattle) and the group (Guyuan and East Asian indicine cattle) and the group (East Asian indicine cattle) and the group (East Asian indicine cattle). Table S4. The frequency of two missense mutations in different cattle breeds. Table S5. Summary of sequencing data.

Acknowledgements

We thank the Key Laboratory of Molecular Cell Breeding of Ruminants in Ningxia Hui Autonomous Region for providing computing resources. We thank the High-Performance Computing (HPC) of Northwest A&F University (NWAFU) for providing computing resources. We thank Dr. Chenglong Qiao (Ningxia University) for providing the geographical information.

Authors' contributions

SL, CZL, NBC and YM conceived and designed the study. SL and XF collected the data. SL, HYY, XYL, YL and NBC performed bioinformatics analysis, and visualized the results. SL, HYY and XF wrote the original draft. SL, CZL, NBC and YM reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [U22A20506, 32472871, and 32341054], the Key Research and Development Program of Ningxia Projects [2024BBF01007, 2023BCF01006], and the earmarked fund for the China Agriculture Research System of MOF and MARA [CARS-37].

Data availability

The newly generated sequences for 39 Guyuan cattle are available from the Sequence Read Archive (SRA) with the Bioproject accession number PRJNA574857. The publicly available sequences were downloaded from the SRA with the following project accession numbers: PRJNA176557, PRJNA256210, and PRJNA176557 (Angus, Hereford, Simmental, and Gelbvieh cattle); PRJNA210523, PRJNA283480, PRJNA379859, PRJNA658727, PRJNA898088, PRJNA598339, and PRJNA823479 (Hanwoo, Yanbian, Xizang, Chaidamu, Anxi, and Mongolian cattle); PRJNA1021039, PRJNA658727, PRJNA83480 (Zaosheng and Qinchuan cattle); PRJNA1021039, PRJNA658727, PRJNA823479, and PRJNA283480 (Xiangnan, Weizhou, and Hainan cattle); PRJNA658727 (Red Sindhi, Dhanni, and Cholistani cattle); PRJNA547460 (Bubalus bubalis).

Declarations

Ethics approval and consent to participate

All experimental procedures involving Guyuan cattle were conducted in accordance with the guidelines set by the Institutional Animal Care and Use Committee of Ningxia University (NXU-A-2023–162) and complied with the Regulations for the Administration of Affairs Concerning Experimental Animals of China.

Consent for publication

Not applicable.

Competing interests

Ningbo Chen is an Editorial Board Member for BMC Biology, but was not involved in the editorial process of this manuscript, including peer review. The manuscript was handled by a member of the journal's editorial team. The authors declare no competing interests.

Received: 16 January 2025 Accepted: 10 April 2025 Published online: 13 May 2025

References

- Taberlet P, Valentini A, Rezaei HR, Naderi S, Pompanon F, Negrini R, Ajmone-Marsan P. Are cattle, sheep, and goats endangered species? Mol Ecol. 2008;17(1):275–84.
- Decker JE, McKay SD, Rolf MM, Kim J, Molina AA, Sonstegard TS, Hanotte O, Götherström A, Seabury CM, Praharani L, et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PloS Genet. 2014;10(3): e1004254.
- 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.
- 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.
- Utsunomiya YT, Milanesi M, Fortes M, Porto-Neto LR, Utsunomiya A, Silva M, Garcia JF, Ajmone-Marsan P. Genomic clues of the evolutionary history of Bos indicus cattle. Anim Genet. 2019;50(6):557–68.
- Kim K, Kim D, Hanotte O, Lee C, Kim H, Jeong C. Inference of admixture origins in indigenous African cattle. Mol Biol Evol. 2023;40(12):msad257.
- Kim K, Kwon T, Dessie T, Yoo D, Mwai OA, Jang J, Sung S, Lee S, Salim B, Jung J, et al. The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. Nat Genet. 2020;52(10):1099–110.
- 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.
- 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.
- Hou J, Guan X, Xia X, Lyu Y, Liu X, Mazei Y, Xie P, Chang F, Zhang X, Chen J, et al. Evolution and legacy of East Asian aurochs. Sci Bull. 2024;69(21):3425–33.
- Ward JA, MacHugh DE. Cattle genomics: aurochs admixture in East Asia. Animal Research and One Health. Published online ahead of print, 12 February 2025. https://doi.org/10.1002/aro2.102.
- Chen N, Huang J, Zulfiqar A, Li R, Xi Y, Zhang M, Dang R, Lan X, Chen H, Ma Y, et al. Population structure and ancestry of Qinchuan cattle. Anim Genet. 2018;49(3):246–8.
- Yu H, Zhang K, Cheng G, Mei C, Wang H, Zan L. Genome-wide analysis reveals genomic diversity and signatures of selection in Qinchuan beef cattle. BMC Genomics. 2024;25(1):558.
- Zhang Y, Wei Z, Zhang M, Wang S, Gao T, Huang H, Zhang T, Cai H, Liu X, Fu T et al. Population structure and selection signal analysis of Nanyang cattle based on whole-genome sequencing data. Genes-Basel. 2024;15(3):351.
- Hu M, Shi L, Yi W, Li F, Yan S. Identification of genomic diversity and selection signatures in Luxi cattle using whole-genome sequencing data. Anim Biosci. 2024;37(3):461–70.
- Xia X, Zhang S, Zhang H, Zhang Z, Chen N, Li Z, Sun H, Liu X, Lyu S, Wang X, et al. Assessing genomic diversity and signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. BMC Genomics. 2021;22(1):43.

- Ma X, Cheng H, Liu Y, Sun L, Chen N, Jiang F, You W, Yang Z, Zhang B, Song E et al. Assessing genomic diversity and selective pressures in Bohai Black cattle using whole-genome sequencing data. Animals- Basel 2022;12(5):665.
- 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.
- Guyuan County Annals Compilation Committee: Guyuan County Annals. Ningxia People's Publishing House, 1993.
- Qin G, Chang H, Geng S, Qiu H, Xu G. Study on blood polymorphism of Guyuan cattle. Chinese Cattle Science. 1997;21(3):13–15.
- Flori L, Thevenon S, Dayo GK, Senou M, Sylla S, Berthier D, Moazami-Goudarzi K, Gautier M. Adaptive admixture in the West African bovine hybrid zone: insight from the Borgou population. Mol Ecol. 2014;23(13):3241–57.
- 22. Muhammad AM, Sharma VK, Pandey S, Kumaresan A, Srinivasan A, Datta TK, Mohanty TK, Yadav S. Identification of biomarker candidates for fertility in spermatozoa of crossbred bulls through comparative proteomics. Theriogenology. 2018;119:43–51.
- Borowska A, Szwaczkowski T, Kamiński S, Hering DM, Kordan W, Lecewicz M. Identification of genome regions determining semen quality in Holstein-Friesian bulls using information theory. Anim Reprod Sci. 2018;192:206–15.
- Ghoreishifar M, Vahedi SM, Salek AS, Khansefid M, Pryce JE. Genome-wide assessment and mapping of inbreeding depression identifies candidate genes associated with semen traits in Holstein bulls. BMC Genomics. 2023;24(1):230.
- Qi H. New twists in humoral immune regulation by SLAM family receptors. J Exp Med. 2021;218(3):e20202300.
- Chen Q, Huang B, Zhan J, Wang J, Qu K, Zhang F, Shen J, Jia P, Ning Q, Zhang J, Chen N. Whole-genome analyses identify loci and selective signals associated with body size in cattle. J Anim Sci. 2020;98(3):skaa068.
- Chen L, Pan L, Zeng Y, Zhu X, You L. CPNE1 regulates myogenesis through the PERK-eIF2α pathway mediated by endoplasmic reticulum stress. Cell Tissue Res. 2023;391(3):545–60.
- Hernandez CA, Gonzales NM, Parker CC, Sokolof G, Vandenbergh DJ, Cheng R, Abney M, Sko A, Douglas A, Palmer AA, et al. Genome-wide associations reveal human-mouse genetic convergence and modifiers of myogenesis, CPNE1 and STC2. Am J Hum Genet. 2019;105(6):1222–36.
- Saha PP, Kumar S, Srivastava S, Sinha D, Pareek G, D'Silva P. The presence of multiple cellular defects associated with a novel G50E iron-sulfur cluster scaffold protein (ISCU) mutation leads to development of mitochondrial myopathy. J Biol Chem. 2014;289(15):10359–77.
- Jones RR, Dimet-Wiley A, Haghani A, Da SF, Brightwell CR, Lim S, Khadgi S, Wen Y, Dungan CM, Brooke RT, et al. A molecular signature defining exercise adaptation with ageing and in vivo partial reprogramming in skeletal muscle. J Physiol-London. 2023;601(4):763–82.
- Anderson DM, Cannavino J, Li H, Anderson KM, Nelson BR, McAnally J, Bezprozvannaya S, Liu Y, Lin W, Liu N, et al. Severe muscle wasting and denervation in mice lacking the RNA-binding protein ZFP106. P Natl Acad Sci USA. 2016;113(31):E4494–503.
- Lee H, Hong Y, Kong G, Lee DH, Kim M, Tran Q, Cho H, Kim C, Park S, Kim SH, et al. Yin Yang 1 is required for PHD finger protein 20-mediated myogenic differentiation in vitro and in vivo. Cell Death Differ. 2020;27(12):3321–36.
- 33. de Las HS, Clark SA, Duijvesteijn N, Gondro C, van der Werf J, Chen Y. Combining information from genome-wide association and multi-tissue gene expression studies to elucidate factors underlying genetic variation for residual feed intake in Australian Angus cattle. BMC Genomics. 2019;20(1):939.
- Uemoto Y, Takeda M, Ogino A, Kurogi K, Ogawa S, Satoh M, Terada F. Genetic and genomic analyses for predicted methane-related traits in Japanese Black steers. Anim Sci J. 2020;91(1): e13383.
- Brunes LC, Baldi F, Lopes FB, Lôbo RB, Espigolan R, Costa M, Stafuzza NB, Magnabosco CU. Weighted single-step genome-wide association study and pathway analyses for feed efficiency traits in Nellore cattle. J Anim Breed Genet. 2021;138(1):23–44.
- Lindholm-Perry AK, Kuehn LA, Smith TP, Ferrell CL, Jenkins TG, Freetly HC, Snelling WM. A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle(1). Anim Genet. 2012;43(2):216–9.

- Arif A, Terenzi F, Potdar AA, Jia J, Sacks J, China A, Halawani D, Vasu K, Li X, Brown JM, et al. EPRS is a critical mTORC1-S6K1 effector that influences adiposity in mice. Nature. 2017;542(7641):357–61.
- Hay EH, Roberts A. Genome-wide association study for carcass traits in a c omposite beef cattle breed. Livest Sci. 2018;213:35–43.
- Kumar G, Thomas B, Mensa-Wilmot K. Pseudokinase NRP1 facilitates endocytosis of transferrin in the African trypanosome. Sci Rep-UK. 2022;12(1):18572.
- Naji MM, Utsunomiya YT, Sölkner J, Rosen BD, Mészáros G. Assessing Bos taurus introgression in the UOA Bos indicus assembly. Genet Sel Evol. 2021;53(1):96.
- Xu L, Zhou K, Huang X, Chen H, Dong H, Chen Q. Whole-genome resequencing provides insights into the diversity and adaptation to desert environment in Xinjiang Mongolian cattle. BMC Genomics. 2024;25(1):176.
- Wei X, Li S, Yan H, Chen S, Li R, Zhang W, Chao S, Guo W, Li W, Ahmed Z, et al. Unraveling genomic diversity and positive selection signatures of Qaidam cattle through whole-genome re-sequencing. Anim Genet. 2024;55(3):362–76.
- Lyu Y, Yao T, Chen Z, Huangfu R, Cheng H, Ma W, Qi X, Li F, Chen N, Lei C. Genomic characterization of dryland adaptation in endangered Anxi cattle in China. Anim Genet. 2024;55(3):352–61.
- 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. 2024;69(21):3415-24.
- 45. Xia X, Zhang F, Li S, Luo X, Peng L, Dong Z, Pausch H, Leonard AS, Crysnanto D, Wang S, et al. Structural variation and introgression from wild populations in East Asian cattle genomes confer adaptation to local environment. Genome Biol. 2023;24(1):211.
- Gondaira S, Nishi K, Fujiki J, Iwano H, Watanabe R, Eguchi A, Hirano Y, Higuchi H, Nagahata H. Innate immune response in bovine neutrophils stimulated with Mycoplasma bovis. Vet Res. 2021;52(1):58.
- Gondaira S, Nishi K, Iwano H, Fujiki J, Watanabe R, Eguchi A, Hirano Y, Higuchi H, Nagahata H. Transcriptome analysis of Mycoplasma bovis stimulated bovine peripheral blood mononuclear cells. Vet Immunol Immunop. 2021;232: 110166.
- Wang N, Halibozek PJ, Yigit B, Zhao H, O'Keeffe MS, Sage P, Sharpe A, Terhorst C. Negative regulation of humoral immunity due to interplay between the SLAMF1, SLAMF5, and SLAMF6 Receptors. Front Immunol. 2015;6:158.
- Sood V, Rodas-González A, Valente TS, Virtuoso M, Li C, Lam S, López-Campos Ó, Segura J, Basarab J, Juárez M. Genome-wide association study for primal cut lean traits in Canadian beef cattle. Meat Sci. 2023;204: 109274.
- Li X, Lian L, Zhang D, Qu L, Yang N. gga-miR-26a targets NEK6 and suppresses Marek's disease lymphoma cell proliferation. Poultry Sci. 2014;93(5):1097–105.
- Perotin JM, Polette M, Deslée G, Dormoy V. CiliOPD: a ciliopathy-associated COPD endotype. Resp Res. 2021;22(1):74.
- 52. Fu B, Xue W, Zhang H, Zhang R, Feldman K, Zhao Q, Zhang S, Shi L, Pavani KC, Nian W, et al. MicroRNA-325-3p facilitates immune escape of mycobacterium tuberculosis through targeting LNX1 via NEK6 accumulation to promote anti-apoptotic STAT3 signaling. MBio. 2020;11(3):10–128.
- Wang R, Long T, Hassan A, Wang J, Sun Y, Xie XS, Li X. Cryo-EM structures of intact V-ATPase from bovine brain. Nat Commun. 2020;11(1):3921.
- 54. Ogawa S, Darhan H, Suzuki K. Genetic and genomic analysis of oxygen consumption in mice. J Anim Breed Genet. 2022;139(5):596–610.
- Olsson AH, Yang BT, Hall E, Taneera J, Salehi A, Nitert MD, Ling C. Decreased expression of genes involved in oxidative phosphorylation in human pancreatic islets from patients with type 2 diabetes. Eur J Endocrinol. 2011;165(4):589–95.
- Chen S, Liu S, Shi S, Jiang Y, Cao M, Tang Y, Li W, Liu J, Fang L, Yu Y, et al. Comparative epigenomics reveals the impact of ruminant-specific regulatory elements on complex traits. BMC Biol. 2022;20(1):273.
- Andrews G, Fan K, Pratt HE, Phalke N, Karlsson EK, Lindblad-Toh K, Gazal S, Moore JE, Weng Z. Mammalian evolution of human cisregulatory elements and transcription factor binding sites. Science. 2023;380(6643):eabn7930.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.

- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
- Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292–4.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del AG, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8.
- 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.
- 64. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7.
- 65. Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. PLoS One. 2014;9(4): e93766.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.
- 67. 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.
- 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.
- Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, Huelsenbeck JP, Ronquist F, Swofford DL, Cummings MP, et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst Biol. 2012;61(1):170–3.
- Maier R, Flegontov P, Flegontova O, Işıldak U, Changmai P, Reich D. On the limits of fitting complex models of population history to f-statistics. Elife. 2023;12:e85492.
- Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet. 2013;93(2):278–88.
- Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. Inferring admixture histories of human populations using linkage disequilibrium. Genetics. 2013;193(4):1233–54.
- Dias-Alves T, Mairal J, Blum M. Loter: a software package to infer local ancestry for a wide range of species. Mol Biol Evol. 2018;35(9):2318–26.
- 74. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80–92.
- Chen N, Fu W, Zhao J, Shen J, Chen Q, Zheng Z, Chen H, Sonstegard TS, Lei C, Jiang Y. BGVD: an integrated database for bovine sequencing variations and selective signatures. Genom Proteom Bioinf. 2020;18(2):186–93.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9.
- Dorji J, Reverter A, Alexandre PA, Chamberlain AJ, Vander-Jagt CJ, Kijas J, Porto-Neto LR. Ancestral alleles defined for 70 million cattle variants using a population-based likelihood ratio test. Genet Sel Evol. 2024;56(1):11.
- Szpiech ZA: selscan 2.0: scanning for sweeps in unphased data. Bioinformatics 2024;40(1):btae006.
- Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics. 2012;28(8):1176–7.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
- Ma R, Kuang R, Zhang J, Sun J, Xu Y, Zhou X, Han Z, Hu M, Wang D, Fu Y et al. Annotation and assessment of functional variants in livestock through epigenomic data. J GENET GENOMICS. Published online ahead of print, 4 April 2025. https://doi.org/10.1016/j.jgg.2025.03.013.

Page 16 of 16

- Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. The Annals of Applied Statistics. 2011;5(3):1752–1779.
- Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, Saelao P, Waters S, Xiang R, Chamberlain A, et al. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. Nat Commun. 2021;12(1):1821.
- Fu W, Wang R, Nanaei HA, Wang J, Hu D, Jiang Y. RGD v2.0: a major update of the ruminant functional and evolutionary genomics database. Nucleic Acids Res. 2022;50(D1):D1091–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.