



## OPEN ACCESS

## EDITED BY

Fei Hao,  
Northumbria University, United Kingdom

## REVIEWED BY

Antoine Fages,  
Université Toulouse III Paul Sabatier, France  
Wei Shi,  
Chinese Academy of Sciences (CAS), China  
Sarah Meirelles,  
Universidade Federal de Lavras, Brazil

## \*CORRESPONDENCE

Kairat Iskhani,  
✉ kayrat\_ishan@mail.ru  
Gemingguli Muhatai,  
✉ gmgl-113@foxmail.com

RECEIVED 27 May 2024

REVISED 26 August 2025

ACCEPTED 26 August 2025

PUBLISHED 12 November 2025

## CITATION

Tang C, Yang B, Dawulietihan G, Xue L, Liu S, Yalimaimaiti Y, Wang Q, Yang N, Sun X, Wang Y, Wumaier A, Khizat S, Assanbayev T, Kozhanov Z, Attokurov K, Obdunov E, Li H, Reheman A, Zhou X, Aizimu W, Iskhani K and Muhatai G (2025) The genetic diversity and population structure of native horse breeds in Xinjiang, China.  
*Front. Genet.* 16:1439312.  
doi: 10.3389/fgene.2025.1439312

## COPYRIGHT

© 2025 Tang, Yang, Dawulietihan, Xue, Liu, Yalimaimaiti, Wang, Yang, Sun, Wang, Wumaier, Khizat, Assanbayev, Kozhanov, Attokurov, Obdunov, Li, Reheman, Zhou, Aizimu, Iskhani and Muhatai. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The genetic diversity and population structure of native horse breeds in Xinjiang, China

Chi Tang<sup>1,2</sup>, Baoyu Yang<sup>1,2</sup>, Gulibaheti Dawulietihan<sup>3</sup>, Li Xue<sup>4</sup>, Shuyuan Liu<sup>5</sup>, Yina mujiang Yalimaimaiti<sup>6</sup>, Qingzheng Wang<sup>7</sup>, Na Yang<sup>1,2,7</sup>, Xiaoyuan Sun<sup>7</sup>, Yaru Wang<sup>7</sup>, Ailifeire Wumaier<sup>7</sup>, Serik Khizat<sup>8</sup>, Tolegen Assanbayev<sup>9</sup>, Zhassulan Kozhanov<sup>10</sup>, Kursantbek Attokurov<sup>11</sup>, Elmurat Obdunov<sup>11</sup>, Hangsen Li<sup>7</sup>, Aikebaier Reheman<sup>7</sup>, Xiaoling Zhou<sup>7</sup>, Wumaierjiang Aizimu<sup>2,7</sup>, Kairat Iskhani<sup>8\*</sup> and Gemingguli Muhatai<sup>1,2,7\*</sup>

<sup>1</sup>College of Life Science and Technology, Tarim University, Alar, Xinjiang, China, <sup>2</sup>Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, Alar, Xinjiang, China, <sup>3</sup>Agricultural (Animal Husbandry) Development Service in Tuerhong Township, Fuyun, Xinjiang, China, <sup>4</sup>Animal Husbandry Workstation of Fuyun County, Fuyun, Xinjiang, China, <sup>5</sup>Animal Husbandry Workstation of Balikun County, Balikun, Xinjiang, China, <sup>6</sup>Animal Husbandry and Veterinary Station of Kalayagazi Town, Yining, Xinjiang, China, <sup>7</sup>Key Laboratory of Tarim Animal Husbandry Science and Technology, Xinjiang Production & Construction Corps, Alar, Xinjiang, China, <sup>8</sup>Physiology, Morphology and Biochemistry, Kazakh National Agrarian Research University, Almaty, Kazakhstan, <sup>9</sup>Zootechnology and Veterinary Medicine, Toraighyrov University, Pavlodar, Kazakhstan, <sup>10</sup>Horse Breeding Department, Kazakh Research Institute of Livestock and Forage Production, Almaty, Kazakhstan, <sup>11</sup>Osh State University, Osh, Kyrgyzstan

**Introduction:** Xinjiang is a region renowned for its rich diversity of native horse breeds, making it one of the most affluent equine genetic resource areas in China. While prized for their high adaptability and tolerance to roughage, the conservation of these native breeds faces challenges from the introduction of external breeds and industrial changes. Furthermore, the unknown population structure of Xinjiang horse breeds has hindered effective conservation efforts.  
**Methods:** This study presents the first comprehensive Single Nucleotide Polymorphism (SNP) analysis of seven Xinjiang native horse breeds. We utilized 10X whole-genome sequencing to assess their genetic diversity, population structure, and genetic relationships.

**Results:** Our findings revealed a high level of population genetic diversity among the Xinjiang native horse breeds. These breeds exhibited significant genetic differentiation from other horse breeds originating from Europe, Central Asia, Western Asia, and other parts of China. Evidence of frequent historical gene flow was detected, particularly among breeds in northern Xinjiang, which were shown to be more closely related to each other.

**Discussion:** This study elucidates the distribution patterns, evolutionary characteristics, and substantial genetic diversity of Xinjiang's native horse breeds. The results provide crucial insights into their unique genetic background and population history. These findings offer valuable theoretical

support for establishing core conservation groups of local germplasm, guiding future breeding programs for new cultivars, and further exploration of the characteristics inherent to Xinjiang's native horse genetic resources.

#### KEYWORDS

native horse breed, whole genome sequencing, genetic diversity, population structure, genetic resources conservation

## 1 Introduction

The domestic horse (*Equus ferus caballus*) was successfully domesticated approximately 5,500 years ago (Librado et al., 2021). However, recent studies have proposed that the modern domestic horse originated in the western Eurasian steppe around 4,200 years ago (Librado et al., 2021). Subsequently, they spread across Eurasia, giving rise to various breeds adapted to diverse geographic and climatic environments. Horse domestication has enhanced human work efficiency and propelled human civilization (Schubert et al., 2014). According to the Food and Agriculture Organization (FAO), there are over 900 horse breeds globally, with 694 native breeds. Among these, 138 are located in Asia and 371 in Europe (<https://www.fao.org/>).

The primary purposes for horse keeping include dairy and meat production, racing, equestrian activities, and maintaining free-range populations. However, agricultural advancements have decreased the overall use of horses, significantly reducing the global horse population. Over the past 2,000 years, domestic horses have lost nearly half of their genetic diversity, with genomic heterozygosity decreasing by approximately 16% in the last 200 years (Fages et al., 2019). Genetic diversity and inbreeding are often connected (Spielman et al., 2004), which can inform the efficacy of breeding conservation and offer insights into the sustainable development and utilization of horse populations. This dramatic decline in the global horse population has severely affected the genetic diversity and number of purebred native breeds. The decreased genetic diversity of domestic horses is primarily due to purebred selection practices, which increase deleterious mutations in domestic horse genomes (Orlando, 2020).

China is known for having a diverse and abundant population of horses, with a rich genetic variety (Ling et al., 2011). Furthermore, Chinese horse breeds include the Northern China (NC), Qinghai-Tibetan (QT), and Southwestern (SW) populations (Liu et al., 2019). The Xinjiang Uygur Autonomous Region (XUAR), a prominent horse-producing area in China, has the largest horse population in the country, reaching 1,107,000 horses as of 2022. This region is known for four native (Kazakh, Yanqi, Kyrgyz, Balikun) and two cultivated (Yili and Yiwu) horse breeds, mainly bred for meat. China is the largest global horse meat producer (FAO statistics, 2022), producing 159,068 tons of horse meat, most of which is produced in Xinjiang. Many horses are underutilized and freely grazed for breeding purposes, with only a small number used for meat or riding (including leisure, competitions, celebrations, etc.). However, the demand for sports events and leisure tourism has changed the use of native horse breeds, with some regions crossbreeding native horses with Thoroughbred and Arabian breeds to achieve better competition results. This practice enhances the genetic diversity and productivity of cultivated horse breeds but significantly threatens the genetic purity of native breeds. Thus, there is a risk of extinction for

native horse breeds with small population sizes due to unorganized selection and the introduction of hybrids.

Sequence analysis of mtDNA from eastern Chinese and European horse populations revealed a higher frequency of haplogroup F and a lower frequency of haplogroup D in eastern Chinese populations (McGahern et al., 2006). In contrast, European populations exhibited the opposite trend, suggesting a form of genetic isolation and differentiation between these two populations (McGahern et al., 2006). This finding was further supported by the mtDNA data from ancient horses in China dating back to 2,000–4,000 years ago (Cai et al., 2009).

Xinjiang, situated in Central Asia, the transitional region between eastern China and Europe, The ancient stone culture in Ili River basin and Boltala River basin in Xinjiang may be related to the early domestication activities (Annie and Dexin, 2020). Xinjiang Kazakh, Yanqi, and Barkun horses exhibit higher Y chromosome diversity than other horse breeds in China, possibly due to the expansive grasslands, traditional animal husbandry practices, and minimal human interference in Xinjiang (Han et al., 2015). Research on the genetic diversity and structure of Xinjiang horse breeds is limited, with most studies focusing on microsatellite markers. However, Whole Genome Sequencing (WGS) has successfully revealed the genetic diversity and structure of various horse breeds, such as the Thoroughbred (Tozaki et al., 2021), Mongolian (Huang et al., 2014), Lichuan, Kazakh (Zhang et al., 2018), and Mavari horses (Jun et al., 2014). Nonetheless, no comprehensive genome-wide study has been used to analyze the genetic information of horse breeds in Xinjiang. There is no molecular data for all the horse breeds in this region.

Therefore, this study investigated the population genetic structure of native horse breeds in Xinjiang. The study utilized WGS technology to gather genetic data from seven native horse breeds (taxa) in Xinjiang, China. The native horse breeds included Kazakh (Altay and Yili region), Yanqi, Kyrgyz, Balikun, and other unexplored breeds, such as the Kunlun (Hotan region, Kunlun Mountains) and Tashkurgan (Tashkurgan Tajik Autonomous County). These horse breeds are divided by the Tianshan Mountains, North of the Tianshan Mountains are the northern horses, which include the Kazakh and Balikun breeds, while the remaining horse breeds are distributed south of the Tianshan Mountains as southern horses. Genetic variations within and between these breeds were analyzed and compared with data from the horse breed sequences in the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>). The study provides the first collection of genetic information from all Xinjiang native horse breeds. These findings will enhance our understanding of the population genetics of native horse breeds and shed light on the history of horse domestication in Xinjiang.

## 2 Materials and methods

### 2.1 Sample collection

In this study, 70 peripheral blood samples were collected from seven horse breeds (taxa), with samples taken from ten horses of each breed. The breeds included Kazakh horse in the Altay region (KA), Kazakh horse in the Yili region (KY), Yanqi (YQ), Balikun (BLK), Kyrgyz (KE), Tashkurgan (TX), and Kunlun (KL) horse. To ensure that the samples were representative of the genetic variation of the breeds, all the samples were collected from different remote pastures, with at least two separate populations sampled for each breed, and we used information provided by the owners to avoid the collection of samples from related horses and exclude parents or progeny. Herders were consulted for background information on the selection and breeding history of their native breeds. Subsequently, the genealogical records were reviewed to ensure that the samples were unrelated but representative of the genetic variation of the breeds. Blood samples from 70 domestic horses were used for Genomic DNA extraction using the TaKaRa MiniBEST Whole Blood Genomic DNA Extraction Kit (Takara Bio, Beijing, China).

The genomic DNA concentration was determined using an INVITROGEN Qubit 4.0 Fluorometer (Invitrogen Inc, CA, United States), and its integrity was assessed through agarose gel electrophoresis. Subsequently, the samples were subjected to whole-genome resequencing on the Illumina HiSeqX Ten platform (Illumina Inc, CA, United States) based on the quality of the whole-genome DNA extraction (the nucleic acid concentrations were all greater than 80 ng/ $\mu$ L, with OD260/280 ratios ranging from 1.8 to 2.0). The genomic DNA samples were stored at  $-80^{\circ}\text{C}$  in a freezer. In addition, sequencing data of 32 horses (including nine domestic horse breeds and one wild horse population) were downloaded from the NCBI (Supplementary Table S1). The ten horse breeds included the Akhal-Teke ( $n = 1$ ), Arabian ( $n = 2$ ), Curly ( $n = 2$ ), Debaos pony ( $n = 5$ ), Mongolian ( $n = 4$ ), Sorraia ( $n = 1$ ), Thoroughbred ( $n = 4$ ), Tibetan ( $n = 5$ ), Yakut ( $n = 6$ ), and Przewalski ( $n = 2$ ).

### 2.2 Detection and quality control of the SNP sites

The whole genome sequences of 70 domestic horse samples yielded approximately 1831.95 GB of raw data (average of 26 GB/sample), with the Q20 and Q30 base quality exceeding 95.00% and 90.00%, respectively (Supplementary Table S1). After removing low-quality reads, each horse genome had a 10-fold coverage. The high-quality sequencing data from 70 horses, along with the sequencing data of 32 horses obtained from NCBI, were mapped to the domestic horse reference genome, EquCab3.0 (Kalbfleisch et al., 2018). Using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2010). Eight samples with low matching rates were removed (Supplementary Table S2). Next, SAMtools v1.3 (Li et al., 2009) and GATK v3.1 (DePristo et al., 2011) were utilized to detect SNPs at variant sites. The 62 equine SNP datasets were merged with the 32 SNP datasets from the NCBI using PLINK v1.9 (Chang et al., 2015). Subsequently, SNP loci with <90% detection rates, a <0.05 minor allele frequency (MAF), and >1% missing genotype rates were filtered out using pi-

hat in PLINK v1.9 (Chang et al., 2015), and the pedigree relationships between samples were evaluated (Petersen et al., 2013). The filter parameters were as follows: (1) Fisher test of strand bias (FS)  $\leq 60$ ; (2) HaplotypeScore  $\leq 13.0$ ; (3) Mapping Quality (MQ)  $\geq 40$ ; (4) Quality Depth (QD)  $\geq 2$ ; (5) ReadPosRankSum  $\geq -8.0$ ; and (6) MQRankSum  $> -12.5$  (Supplementary Table S3; Supplementary Figures S1–S3). Linkage disequilibrium (LD) was evaluated by calculating  $r^2$  at the genome-wide intervals of 25 SNPs with a window size of 100 SNPs. The dataset was further filtered to remove samples with low quantities, retaining others representative of the five horse breeds, using a 0.2 threshold for judging LD (--indep-pairwise 100 25 0.2).

### 2.3 Analysis of genetic diversity and population structure

The expected heterozygosity ( $H_e$ ) and coefficient of inbreeding ( $F_{is}$ ) were calculated using the PLINK software. The  $H_e$  for each population was performed through the Hardy-Weinberg test, while  $F_{is}$  was calculated per sample using the number of pure heterozygotes. Furthermore, the principal component analysis (PCA) and population differentiation index ( $F_{st}$ ) were determined using smartPCA in the EIGENSOFT software v4.2 (Price et al., 2006). The proportion of missing data to be  $\leq 10\%$  was allowed. Phylogenetic trees were constructed using the neighbor-joining (NJ) method in FastTree v2.1.11 (Price et al., 2009). The resulting tree file was subsequently uploaded to the iTol v5 (Letunic and Bork, 2021) for visualization purposes. Phylogenetic networks were visualized in R version 3.5.1. The genetic structure of the population was assessed with the fastSTRUCTURE package (Raj et al., 2014). The results of the fastSTRUCTURE analysis were then visualized using the DISTRUCT software (Rosenberg et al., 2002).

## 3 Results

The four main Xinjiang horse breeds, developed in diverse environmental conditions, exhibited distinct characteristics (Figure 1). The 94 samples that met the specified quality control criteria generated 26,539,717 SNPs (Supplementary Table S4).

### 3.1 Analysis of genetic diversity

Table 1 shows that the heterozygosity values across all the breeds and populations were notably similar. The observed heterozygosity ( $H_{obs}$ ) reflects the actual proportion of heterozygosity within the population. The genetic diversity of the Xinjiang native horse breeds was found to be relatively low compared to Mongolian and Tibetan horse (Supplementary Table S5). Furthermore, the  $H_{obs}$  for all the populations was lower than the expected heterozygosity ( $H_{exp}$ ), suggesting the presence of varying degrees of inbreeding or genetic drift. Additionally, Fisher's inbreeding coefficient ( $F_{is}$ ) provided further insights into the degree of inbreeding; notably, the YQ population exhibited the highest  $F_{is}$  value (0.1844). The  $F_{is}$  results for the Xinjiang native horse breeds indicated generally low levels of inbreeding, with the lowest coefficients observed in the KE

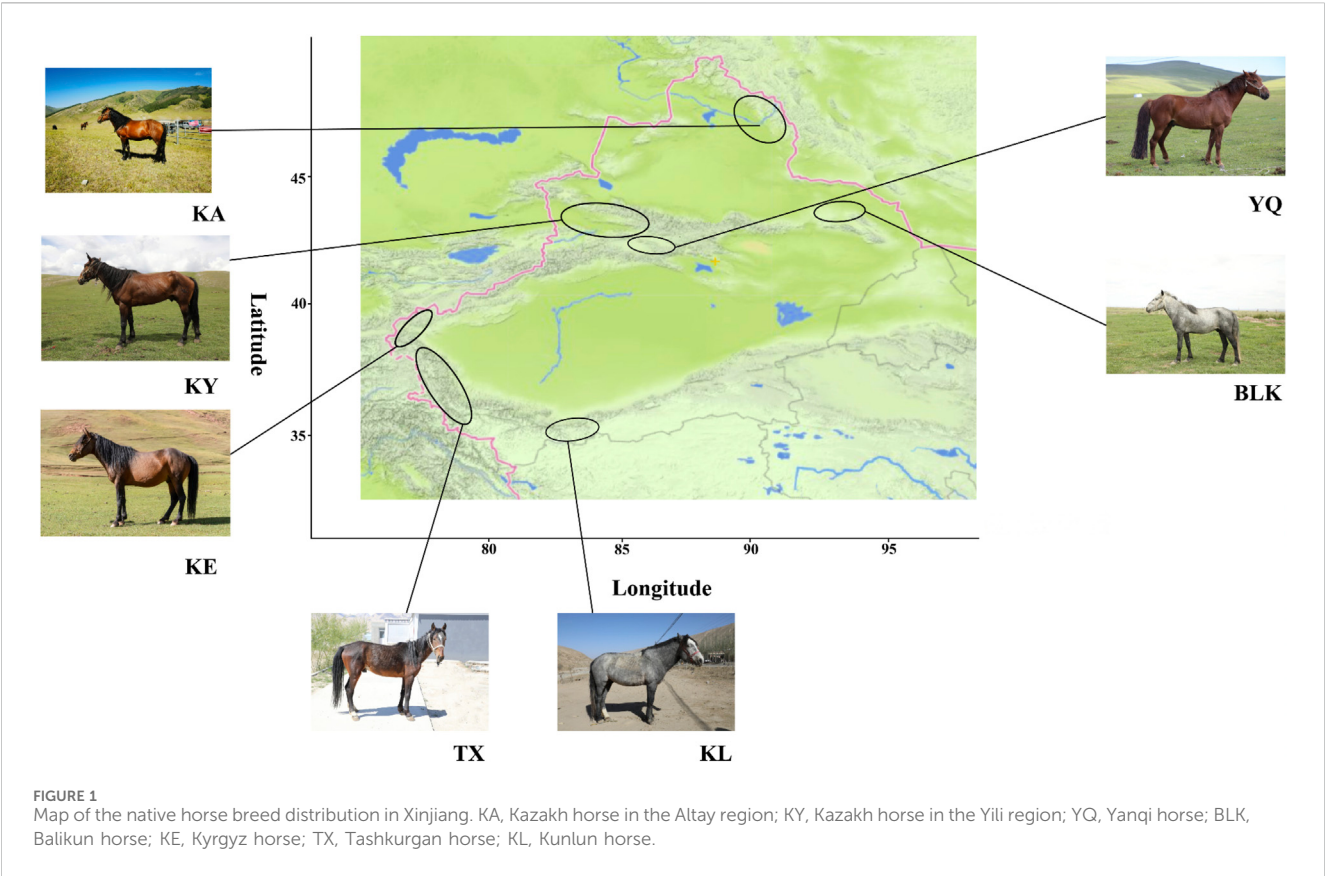
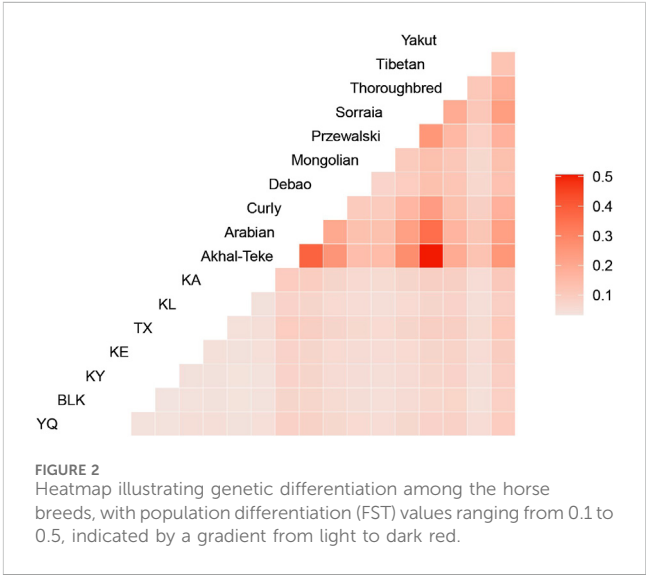


TABLE 1 Comparison of genetic diversity of Xinjiang horse breeds.

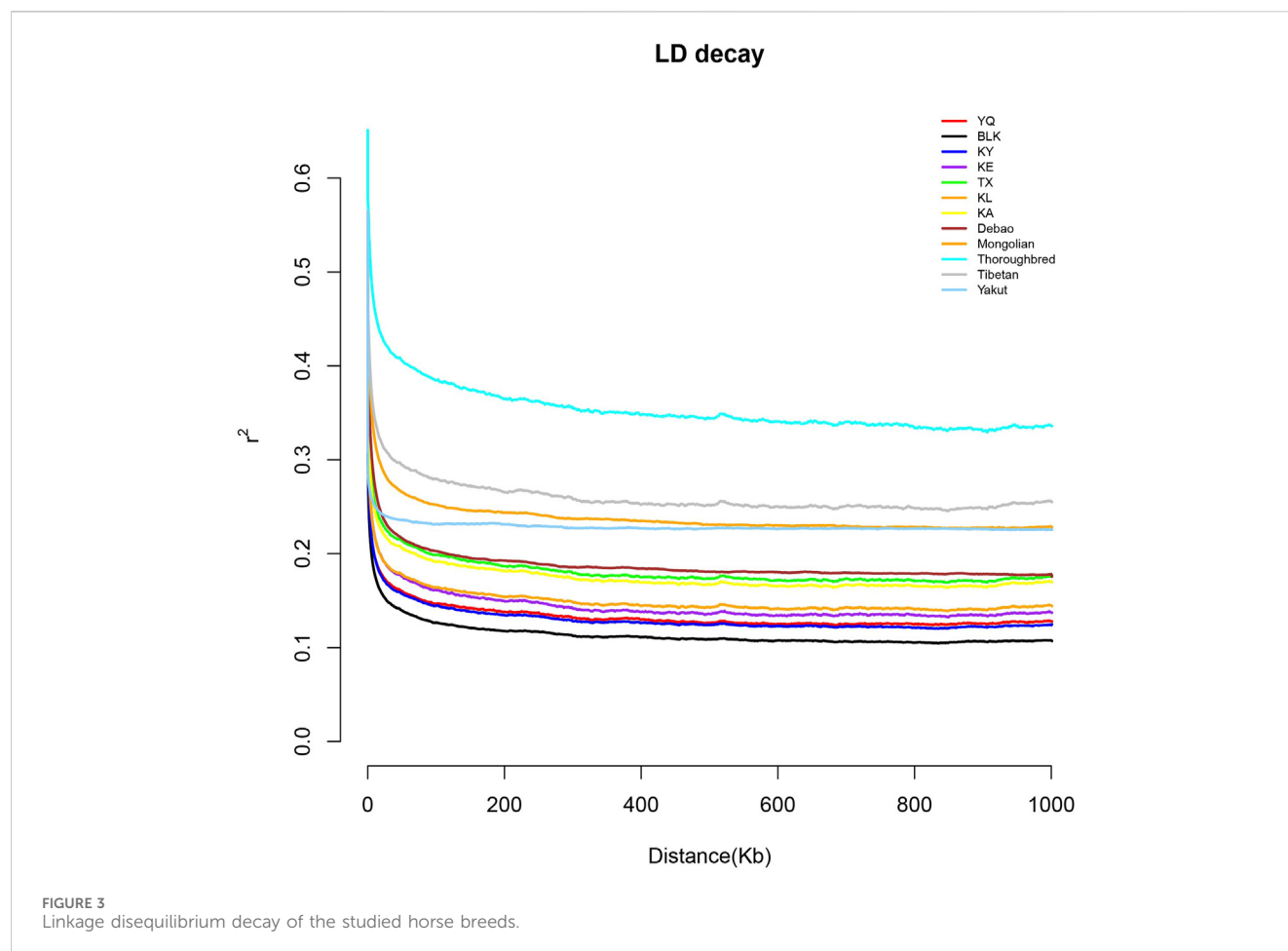
| Breed/Population | Obs Het | Exp Het | Pi     | Fis    |
|------------------|---------|---------|--------|--------|
| YQ               | 0.1940  | 0.2481  | 0.2632 | 0.1844 |
| BLK              | 0.2207  | 0.2541  | 0.2682 | 0.1336 |
| KY               | 0.2330  | 0.2543  | 0.2697 | 0.1013 |
| KE               | 0.2294  | 0.2486  | 0.2619 | 0.0915 |
| TX               | 0.2248  | 0.2472  | 0.2668 | 0.1061 |
| KL               | 0.2408  | 0.2536  | 0.2689 | 0.0781 |
| KA               | 0.2285  | 0.2490  | 0.2693 | 0.1026 |

and KL populations, likely attributed to differences in stallion distributions within these groups. The polymorphic information content (PIC/Pi) value of 0.2689 indicates that the KL population exhibits higher levels of polymorphism and more pronounced genetic differences among individuals. This elevated Pi value reflects a greater genetic diversity, which is crucial for the adaptation of the assessment population. Understanding this genetic diversity is significant for breeding strategies and overall population management.

The genetic differentiation index ( $F_{st}$ ) between the horse groups demonstrates the genetic divergence among the different horse populations. The differentiation within the Xinjiang native horse breeds was relatively lower than that of the other breeds. The  $F_{st}$  values for the Xinjiang horse breeds did not surpass 0.01, indicating



minimal differentiation among the samples. In contrast, the Arabian and Akhal-Teke horses exhibited higher differentiation from the remaining groups, with the highest  $F_{st}$  compared with the other horse breeds (Figure 2). Although Przewalski's horses diverged from domestic horses approximately 40,000 years ago, their  $F_{st}$  values with domestic breeds were lower than expected (Figure 2). This pattern may reflect shared ancestral polymorphisms at conserved neutral loci, as captured by the SNP array. Additionally,



demographic histories, such as bottlenecks during domestication, may have contributed to the homogenization of diversity at certain loci. This may be related to their breeding strategies. Both horse breeds have undergone long-term closed breeding and intensive selection to preserve their unique and exceptional traits, such as the endurance and elegance of Arabian horses, as well as the speed and endurance of Akhal-Teke horses. While these practices maintain the desired breeding characteristics, they also result in a decrease in genetic diversity and an increase in interpopulation differentiation.

The Xinjiang native horse breeds had consistently lower LD Decay rates than the domestic native and foreign breeds (Figure 3), suggesting a higher genetic diversity than non-Chinese breeds retrieved from NCBI, with BLK having the highest value in this study. In contrast, Thoroughbreds, having undergone extensive and systematic selective breeding and controlled reproduction to meet specific performance requirements, especially for racing, had the lowest genetic diversity.

### 3.2 Genetic structure

The PCA revealed limited differentiation among the Xinjiang native horse breeds. These native breeds clustered together with considerable overlap (Figure 4). The results of the Principal Component Analysis (PCA) presented in Figure 4A demonstrate

a distinct separation between the native horse breeds of China and the foreign horse breeds, utilizing Przewalski's horse as the reference population. Among them, the Yakut, Thoroughbred, Sorraia, and Curly horse breeds were obviously different from Xinjiang native horse breeds, while Thoroughbred and Yakut showing notable separation from the other horse breeds. Figure 4B shows the distribution of these Xinjiang native horse breeds on principal component PC1 and PC2, with different colored ellipses indicating the 95% confidence intervals for each population. The size and shape of these ellipses reflect the genetic differences between the populations. Among them, the confidence ellipses of KA and KE were smaller and overlapping, indicating that they had a similar genetic background. The Xinjiang native horse breeds are closest in genetic background to Mongolian and Tibetan horses, possibly due to the historical exchanges between these breeds. Furthermore, the PCA results revealed significant genetic differentiation between the Central and Western Asian breeds, including Arabian and Akhal-Teke horses, corroborating the  $F_{st}$  results.

The native horses of Xinjiang were classified into a single large taxon, indicating that these horses are largely distinct from non-Xinjiang breeds. This pattern is more likely due to the long-term independent evolutionary history of Xinjiang horses, which has led to the accumulation of unique mutations over generations, rather than solely due to unique genetic characteristics during early domestication. (Figure 5A). Moreover, the Debao horse, which

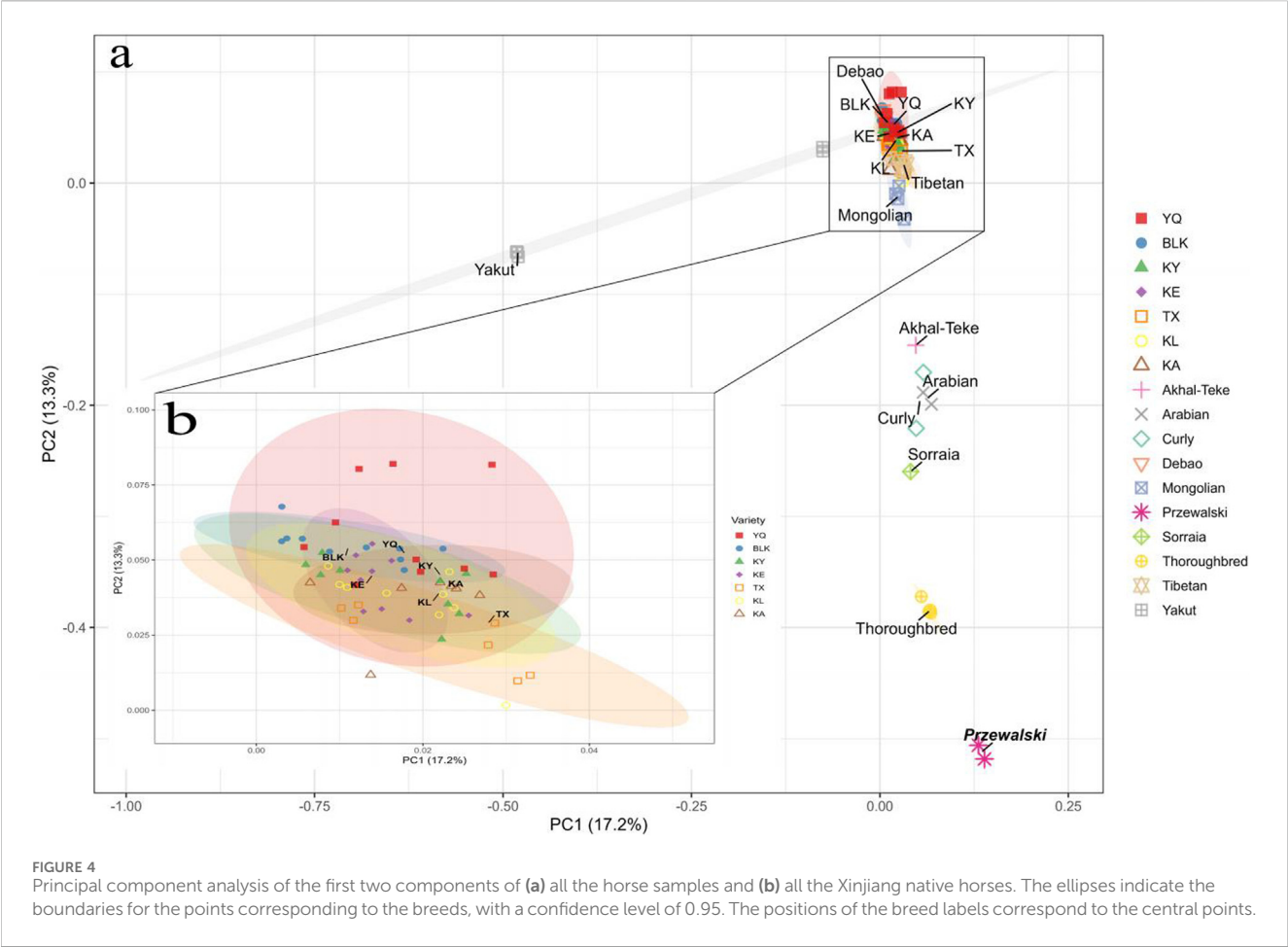


FIGURE 4 Principal component analysis of the first two components of (a) all the horse samples and (b) all the Xinjiang native horses. The ellipses indicate the boundaries for the points corresponding to the breeds, with a confidence level of 0.95. The positions of the breed labels correspond to the central points.

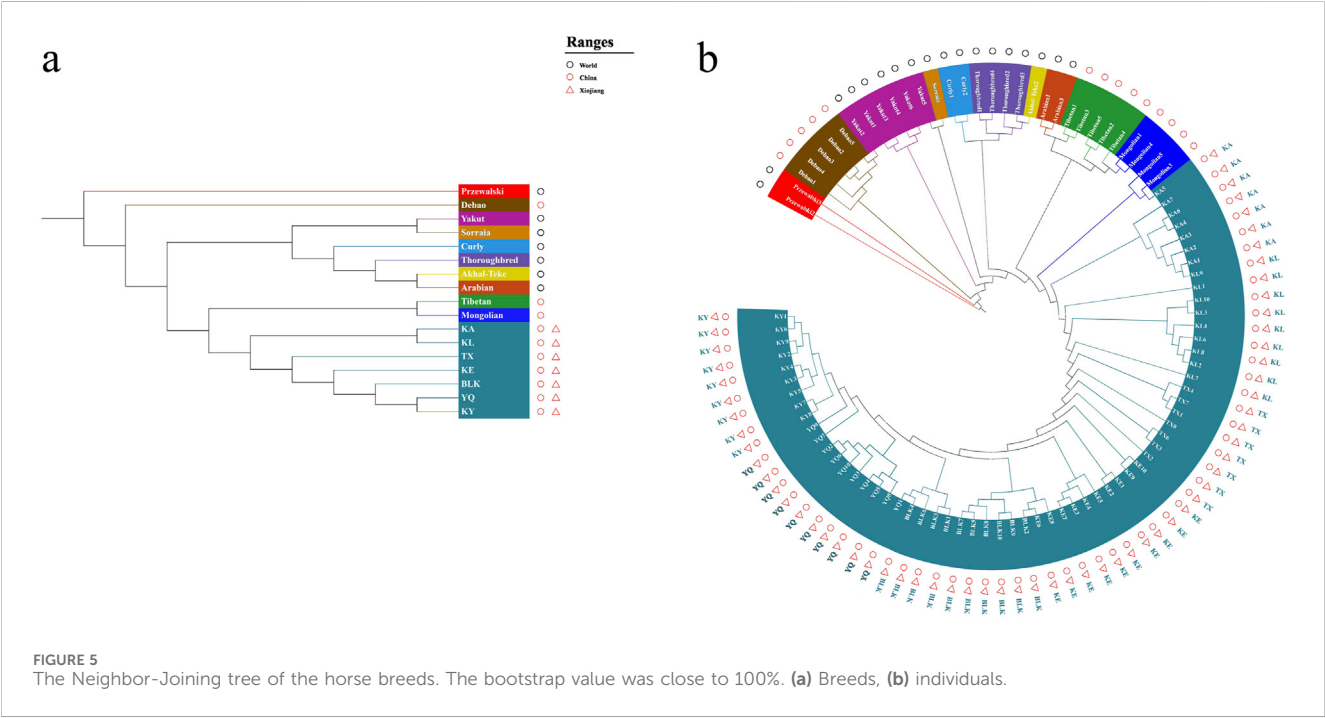


FIGURE 5 The Neighbor-Joining tree of the horse breeds. The bootstrap value was close to 100%. (a) Breeds, (b) individuals.

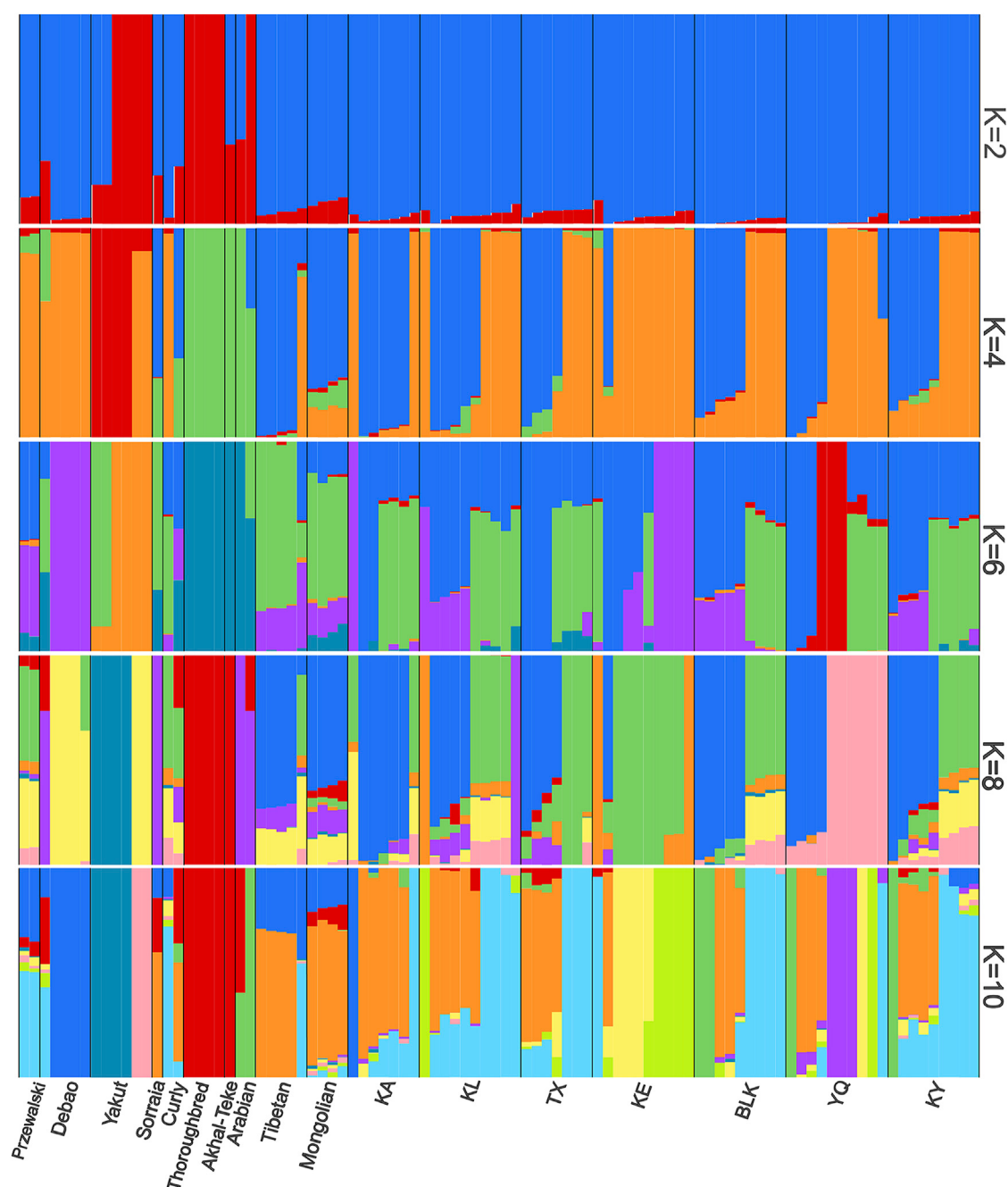


FIGURE 6

Comparative structure of the horse populations with the number of ancestral clusters (K) ranging from 2 to 10. Each bar represents an individual for each breed.

exhibits a unique evolutionary pattern characterized by a distinct genetic lineage, stands out from other breeds. Meanwhile, the remaining Chinese and foreign horse breeds can be broadly categorized into two major genetic clusters. Subsequent classification identified two main groups in southern and northern Xinjiang, highlighting a significant genetic distance between the breeds in these regions and reflecting the geographic distance (Figure 5B). The Debao pony exhibited the most primitive branching, while the Mongolian and Tibetan horses displayed a closer phylogenetic relationship. In Xinjiang, KA, KY, YQ, and BLK

were closely related to the northern Xinjiang breeds, whereas KE and TX were more closely related to the southern Xinjiang breeds, which is consistent with the findings of the PCA results.

The population structure was analyzed for all the horse samples, considering a K value between 2 and 10, representing the ancestral groups (Figure 6). At a value of  $K = 4$ , we could distinguish the separate groups of horse breeds in China and horse breeds in other countries (Supplementary Figure S4). The population genetic structure analysis revealed that at  $K = 6$ , the Xinjiang native horse breeds were distinct. The analysis of  $K = 10$  reveals that

the genetic structure of the native horse breed in Xinjiang is highly complex. This complexity arises not only from random genetic drift or mutations but is more likely attributable to prolonged gene flow.

## 4 Discussion

This study included seven Xinjiang native (KY, YQ, BLK, KE, TX, and KL), three Chinese (Mongolian, Tibetan, and Debao pony), and six foreign (Thoroughbred, Curly, Yakut, Sorraia, Arabian, and Akhal-Teke) horse breeds. Xinjiang native horse breeds exhibited typical genetic characteristics that low inbreeding levels, aligning with the traditional open breeding practices on Xinjiang grasslands. This result suggests strong resistance to inbreeding within this population. Furthermore, Thoroughbreds had the lowest genetic diversity due to varying degrees of bottleneck effects caused by closed artificial selection and high levels of inbreeding (Khanshour et al., 2013; Cosgrove et al., 2020). Ethnic groups in Xinjiang prefer specific horse coat colors; for example, Kazakhs favor dark colors (red bay and black bay), and Tajiks favor light color horses. Therefore, selections are based on coat color preferences. Hence, Thoroughbred and Arabian crosses with desired coat colors were introduced. Interestingly, the genetic structure, which was influenced by coat color preferences, has maintained the genetic diversity of the horse breeds (Druml et al., 2009).

Xinjiang horse breeds are genetically distinct from other Chinese, Central Asian, and European horse breeds. This differentiation is likely primarily driven by the long-term independent evolutionary history following the divergence of these horse breeds from other populations. Additionally, other factors such as historical population bottleneck effects, genetic drift, special environmental adaptability, and targeted artificial selection for specific traits (Castaneda et al., 2019) may have further contributed to their genetic distinctiveness. Moreover, Xinjiang horse breeds have a relatively independent genetic structure and are more closely related to Tibetan and Mongolian horses. The genetic makeup of Xinjiang horse breeds, particularly Kazakh horses, aligns with previous research findings (Liu et al., 2019). Crossbreeding practices in Xinjiang involve using native horse breeds as dams and Arabian or Thoroughbred horses as sires to produce offspring with enhanced speed and endurance for competitive events. Thus, this study included all purebred native breeds, suggesting that other breeds were introduced in the past for crossbreeding to improve performance, albeit with challenges in tracing their pedigree and genealogical information due to insufficient records. This study, based on a genetic analysis of 32 horses from outside Xinjiang using genome-wide autosomal SNPs, revealed significant genetic distinctions, reflecting their geographic origin and breed history (Petersen et al., 2013). Conversely, studies on the Kazakh horse mtDNA indicated that native horse breeds in western China have multiple matrilineal origins and exchange genetic material with each other and other horse breeds during mobility (Zhang et al., 2012). The PCA and structural results from this study indicate that Kazakhstan possesses a complex genetic background, with both autosomal and mitochondrial DNA (Gemingguli et al., 2016) findings corroborating each other.

Historical interbreeding probably influenced horse breeds divergence. Archaeological data revealed significant gene flow between populations in the Pamir Plateau and Ferghana Valley, acting as a 'transmitter' for exchanges between eastern (Tarim Basin, China) and western (Bactria, Uzbekistan) populations (Hemphill and Mallory, 2004). Horse breeds in the Middle East are closely related to local horse breeds in Xinjiang and may be influenced by human activities. Future research should investigate the gene flow between these populations in conjunction with historical events.

In our analysis, we observed distinct genetic clusters among different horse breeds, which highlights the genetic differentiation between breeds. However, it is important to note that within each breed, there may exist multiple distinct subgroups with significant genetic differences. This intrabreed heterogeneity could be due to factors such as regional differences, selective breeding practices, or historical gene flow events. Genetic differentiation has been observed within various native horse breeds in Xinjiang, particularly in the Northern Xinjiang region. This differentiation is primarily attributed to the introduction of foreign germplasm for crossbreeding. Since herders often introduced Thoroughbreds for crossbreeding, this could seriously affect the purebred germplasm resources and reduce the genetic diversity of the native horse population in Xinjiang.

## 5 Conclusions

This study presents the first comprehensive analysis of whole genome sequences of Xinjiang native horse breeds, revealing the genetic diversity and structure of these breeds. Xinjiang native horse breeds exhibited higher genetic diversity, with evidence of gene exchange and similarities in the genetic backgrounds of the different groups. The genetic structure of the southern and northern border breeds differed. Additionally, Tashkurgan and Kunlun horses may represent new cryptic horse breeds. Phylogenetic analysis shows that Tashkurgan and Kunlun horses form a distinct clade separate from other Xinjiang breeds (see Figure 5). These horses also exhibit unique genetic signatures in their SNP profiles, indicating a degree of genetic isolation and differentiation. Further investigation is warranted to fully characterize these potential cryptic breeds. Our findings assessed genetic variability and inbreeding within these horse breeds, thereby laying a foundation for future horse breeding policies. They could inform the delineation of protected areas and breeding grounds to prevent large-scale inbreeding and hybridization and will provide a valuable reference for the conservation and utilization of native horse breeds in Xinjiang.

## Data availability statement

The data presented in the study are deposited in the NCBI SRA repository, accession number PRJNA1185891.

## Ethics statement

The animal studies were approved by Institutional Animal Care and Use Committee at Tarim University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

## Author contributions

CT: Writing – original draft, Writing – review and editing. BY: Writing – original draft. GD: Writing – original draft. LX: Resources, Supervision, Writing – review and editing. SL: Resources, Supervision, Writing – review and editing. YY: Resources, Supervision, Writing – review and editing. QW: Data curation, Investigation, Writing – review and editing. NY: Data curation, Investigation, Writing – review and editing. XS: Formal Analysis, Writing – review and editing. YW: Formal Analysis, Writing – review and editing. AW: Resources, Writing – review and editing. SK: Methodology, Writing – review and editing. TA: Methodology, Writing – review and editing. ZK: Methodology, Writing – review and editing. KA: Methodology, Writing – review and editing. EO: Methodology, Writing – review and editing. HL: Resources, Writing – review and editing. AR: Formal Analysis, Writing – review and editing. XZ: Conceptualization, Writing – review and editing. WA: Methodology, Writing – review and editing. KI: Conceptualization, Project administration, Supervision, Writing – review and editing. GM: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China [31960643, 31560616]; Graduate Research Innovation Project of Tarim University [TDBSCX202210]; Open project of Xinjiang

Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin [BRZD2203]; Agricultural germplasm resources survey, collection, protection and identification service project of the Ministry of Agriculture and Rural Affairs [19221122].

## Acknowledgments

The authors gratefully thanks to the Agriculture and Rural Bureau of Fuyun, Nileke, Barikun, Hotan, Tashkurgan, Wuqia, Hejing County, etc. for their support. The computations in this paper were run on the bioinformatics computing platform of Tarim University. We are also grateful to Prof. Indira Beishova (Zhangir Khan West-Kazakhstan Agrarian-Technical University), Prof. Malika Shamekova (Institute of Plant Biology and Biotechnology), and Prof. Abdugani Abdurasulov (Osh State University) for their constructive comments on the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1439312/full#supplementary-material>

## References

- Annie, C., and Dexin, C. (2020). Results of field research on ancient stonework in the river valleys of bortala and ili in Western tian Shan (xinjiang, China). *Asian Perspect.* 59, 385–420. doi:10.1353/asi.2020.0019
- Cai, D., Tang, Z., Han, L., Speller, C. F., Yang, D. Y., Ma, X., et al. (2009). Ancient DNA provides new insights into the origin of the Chinese domestic horse. *J. Archaeol. Sci.* 36, 835–842. doi:10.1016/j.jas.2008.11.006
- Castaneda, C., Juras, R., Khanshour, A., Randalht, I., Wallner, B., Rigler, D., et al. (2019). Population genetic analysis of the Estonian native horse suggests diverse and distinct genetics, ancient origin and contribution from unique patriline. *Genes (Basel)* 10, 629. doi:10.3390/genes10080629
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8
- Cosgrove, E. J., Sadeghi, R., Schlamp, F., Holl, H. M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S. R., et al. (2020). genome diversity and the origin of the arabian horse. *Sci. Rep.* 10, 9702. doi:10.1038/s41598-020-66232-1
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi:10.1038/ng.806
- Druml, T., Baumung, R., and Sölkner, J. (2009). Pedigree analysis in the Austrian noriker draught horse: genetic diversity and the impact of breeding for coat colour on population structure. *J. Anim. Breed. Genet.* 126, 348–356. doi:10.1111/j.1439-0388.2008.00790.x
- Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., et al. (2019). tracking five millennia of horse management with extensive ancient genome time series. *Cell* 177, 1419–35.e31. doi:10.1016/j.cell.2019.03.049
- Gemingguli, M., Iskhan, K. R., Li, Y., Qi, A., Wunirifu, W., Ding, L. Y., et al. (2016). Genetic diversity and population structure of Kazakh horses (*Equus caballus*) inferred from mtDNA sequences. *Genet. Mol. Res.* 15. doi:10.4238/gmr.15048618
- Han, H., Zhang, Q., Gao, K., Yue, X., Zhang, T., Dang, R., et al. (2015). Y-Single nucleotide polymorphisms diversity in Chinese Indigenous horse. *Asian-Australas J. Anim. Sci.* 28, 1066–1074. doi:10.5713/ajas.14.0784

- Hemphill, B. E., and Mallory, J. P. (2004). Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from Western central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. *Am. J. Phys. Anthropol.* 124, 199–222. doi:10.1002/ajpa.10354
- Huang, J., Zhao, Y., Shiraigol, W., Li, B., Bai, D., Ye, W., et al. (2014). Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Sci. Rep.* 4, 4958. doi:10.1038/srep04958
- Jun, J., Cho, Y. S., Hu, H., Kim, H. M., Jho, S., Gadhvi, P., et al. (2014). Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. *BMC Genomics* 15 (Suppl. 9), S4. doi:10.1186/1471-2164-15-S9-S4
- Kalbfleisch, T. S., Rice, E. S., DePriest, M. S., Jr., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., et al. (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun. Biol.* 1, 197. doi:10.1038/s42003-018-0199-z
- Khanshour, A., Conant, E., Juras, R., and Cothran, E. G. (2013). Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. *J. Hered.* 104, 386–398. doi:10.1093/jhered/est003
- Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi:10.1093/nar/gkab301
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi:10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Librado, P., Khan, N., Fages, A., Kusliy, M. A., Suchan, T., Tonasso-Calvière, L., et al. (2021). The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* 598, 634–640. doi:10.1038/s41586-021-04018-9
- Ling, Y. H., Ma, Y. H., Guan, W. J., Cheng, Y. J., Wang, Y. P., Han, J. L., et al. (2011). Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers. *Anim. Genet.* 42, 56–65. doi:10.1111/j.1365-2052.2010.02067.x
- Liu, X., Zhang, Y., Li, Y., Pan, J., Wang, D., Chen, W., et al. (2019). EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol. Biol. Evol.* 36, 2591–2603. doi:10.1093/molbev/msz158
- Mcgahern, A., Bower, M. A. M., Edwards, C. J., Brophy, P. O., Sulimova, G., Zakharov, I., et al. (2006). Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. *Anim. Genet.* 37, 494–497. doi:10.1111/j.1365-2052.2006.01495.x
- Orlando, L. (2020). The evolutionary and historical foundation of the modern horse: lessons from Ancient genomics. *Annu. Rev. Genet.* 54, 563–581. doi:10.1146/annurev-genet-021920-011805
- Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., et al. (2013). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One* 8, e54997. doi:10.1371/journal.pone.0054997
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi:10.1038/ng1847
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi:10.1093/molbev/msp077
- Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi:10.1534/genetics.114.164350
- Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., et al. (2002). Genetic structure of human populations. *Science* 298, 2381–2385. doi:10.1126/science.1078311
- Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., et al. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5661–E5669. doi:10.1073/pnas.1416991111
- Spielman, D., Brook, B. W., and Richard, F. (2004). Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15261–15264. doi:10.1073/pnas.0403809101
- Tozaki, T., Ohnuma, A., Kikuchi, M., Ishige, T., Kakoi, H., Hirota, K. I., et al. (2021). Rare and common variant discovery by whole-genome sequencing of 101 Thoroughbred racehorses. *Sci. Rep.* 11, 16057. doi:10.1038/s41598-021-95669-1
- Zhang, T., Lu, H., Chen, C., Jiang, H., and Wu, S. (2012). Genetic diversity of mtDNA D-loop and maternal origin of three Chinese native horse breeds. *Asian-Australas J. Anim. Sci.* 25, 921–926. doi:10.5713/ajas.2011.11483
- Zhang, C., Ni, P., Ahmad, H. I., Gemingguli, M., Baizilaitibei, A., Gulibaheti, D., et al. (2018). Detecting the population structure and scanning for signatures of selection in horses (*Equus caballus*) from whole-genome sequencing data. *Evol. Bioinform Online* 14, 1176934318775106. doi:10.1177/1176934318775106