RESEARCH



Comprehensive genomic analysis of genetic diversity, body size, and origins of the Hetian Gray donkey

Yuan-Yuan Li^{1,2,3}, Yu Liu^{1,2,3}, Shu-Qin Liu^{4,5}, Yu-Gang He¹, Xiao-Ran Lin⁶, Ying Li⁷, Tao Yang^{1,2,3}, Mo Feng^{1,2,3}, He-Tong Zhang^{1,2,3}, Xin-Yu Wang^{1,2,3}, Ru-Yu Yao^{1,2,3}, Meng-Han Zhou^{1,2,3} and Chun-Jiang Zhao^{1,2,3*}

Abstract

Background The Hetian Gray donkey, characterized by its large stature and gray coat color, is a relatively recently identified breed. This breed plays an important role in local agricultural practices and cultural heritage. However, with the widespread adoption of mechanization, its use as a draft animal has gradually diminished. Additionally, the insufficient development of its meat and milk products has contributed to a continuous decline in its population, making its conservation increasingly urgent. Despite its significance, research on the origin and ancestral composition of the Hetian Gray donkey remains limited, which is essential for developing effective conservation strategies. To address this, whole-genome resequencing data of 208 donkeys from seven Chinese indigenous donkey breeds, *Equus africanus somaliensis*, and *Equus kiang* were analyzed to investigate the relationships between the Hetian Gray donkey and other breeds in the present study. The findings of this study will provide valuable scientific evidence for preserving the genetic diversity and unique characteristics of the Hetian Gray donkey.

Results The analysis revealed that wild donkeys had no significant impact on the formation of the Hetian Gray donkey. Among the breeds studied, the Xinjiang donkey and the Guanzhong donkey exhibit the closest genetic affinity with the Hetian Gray donkeys. Additionally, genome-wide selection signature analysis between the Hetian Gray donkey and the Xinjiang donkey identified positive selection for genes associated with body size traits.

Conclusion In summary, the results confirm that geographical distance plays a crucial role in determining the genetic relationships among the Hetian Gray donkey and other populations. The Xinjiang donkey and the Guanzhong donkey played an important role in the formation of the Hetian Gray donkey, and genes related to body size were positively selected during development of the breed.

Keywords Hetian Gray donkey, Genome resequencing, Origin, Population genetic structure, Selection signature

*Correspondence: Chun-Jiang Zhao

cjzhao@cau.edu.cn ¹College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

²State Key Laboratory of Animal Biotech Breeding, Beijing 100193, China



 ³Equine Center, China Agricultural University, Beijing 100193, China
⁴College of Animal Science and Technology, Qingdao Agricultural University, Qingdao 266109, China
⁵Shandong Equine Genetic Resource Gene Bank, Qingdao 266109, China
⁶Beijing General Station of Animal Husbandry, Beijing 100107, China
⁷University of Chinese Academy of Sciences, Beijing 100049, 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 are provided in the article's Creative Commons licence, unless indicate 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

China has a long history of donkey breeding that spans 5,000 years, and is one of the countries with the richest genetic resources in donkeys [1]. There were approximately 1.46 million indigenous donkeys in China in 2023, with more than 150 thousand of them distributed in the Xinjiang Uyghur Autonomous Region (http://www.sta ts.gov.cn), which accounted for 2.76% and 0.28% of the total donkey population in the world (around 52.95 million) (http://www.fao.org), respectively. Xinjiang is a tra ditional and prominent donkey-raising region within the country. The majority of donkeys in the area are classified as Xinjiang donkeys, while a small proportion belongs to the Hetian Gray donkey. The Hetian Gray donkey is an ancient and distinctive indigenous breed mainly distributed in Pishan County in the Hetian region of Xinjiang [2], which were identified and named after the prefecture where they are distributed in 2006 [3]. This breed is recognized for its large size, gray coat color as well as its strong adaptation to the local dry and hot climate [4], and it is distinctive from the most donkeys in Xinjiang, which are characterized by their small body size. Though the Hetian Gray donkey is regarded as a superior donkey breed, its origin still largely remains unknown. As Hetian region is located in the ancient route of the Silk Route, and there have been wild donkeys inhabiting in the South of Xinjiang, which overlap the distribution area of the Hetian Gray donkey, it led to the speculation that the donkey breeds along the Silk Route or the wild donkey may contribute to the breed formation of the Hetian Gray donkey. Affected by agricultural mechanization, the population of the Hetian Gray donkey has decreased in the past decades. Despite the conservation of the breed is supported by the local government, efficient actions are still needed for preserving the Hetian Gray donkey. A study on the origin of Hetian Gray donkeys will provide meaningful information in support of the conservation and utilization of the Hetian Gray donkey. Despite cultural and scientific interest in the Hetian Gray donkey, few efforts have been made to better understand the origin of this unique Chinese indigenous donkey breed (CDB).

In recent years, genome resequencing technology has been widely employed in population genetics studies of various species, and led to significant progresses in the studies on genetic diversity and evolution [5–12]. For donkeys, studies with whole-genome sequencing data showed that Chinese donkeys exhibit high genetic diversity [13–16] and they are the offsprings of the donkeys originally domesticated in Africa [17]. These findings provide a scientific basis for the genetic study and improvement of Chinese indigenous donkey breeds (CDBs). In this study, we conducted whole-genome resequencing on seven CDBs, and analyzed 206 indigenous donkey genomes to explore the genetic relationships between the Hetian Gray donkey and other donkey breeds, providing insights into the origin of the Hetian Gray donkey. The findings of this study offer a meaningful basis for the conservation of this special indigenous donkey breed.

Materials and methods

Sample collection and datasets

In this study, a total of 206 blood samples were randomly collected from seven indigenous donkey breeds in China. The samples included 30 Hetian Gray donkeys and 16 Xinjiang donkeys from the Xinjiang Uygur Autonomous Region, 35 Dezhou donkeys from Shandong Province, 35 Guanzhong donkeys from Shaanxi Province, 30 Guangling donkeys from Shanxi Province, 30 Huaibei Gray donkeys from Anhui Province, and 30 Yunnan donkeys from Yunnan Province. The studied populations were sampled in their places of origin. The Yunnan donkeys were sampled from the local private farms, and all other samples were collected from donkey preserving farms owned by the local governments in the regions where the sampled breeds distributed. Close blood relationships among the sampled animals were avoided. Blood samples were collected from the jugular vein and immediately stored in EDTA anticoagulant tubes. The indigenous donkey breeds sampled in this study are representatives of the main donkey breeding regions across China (Fig. 1).

Genomic DNA was extracted using the Rapid Blood Genomic DNA Extraction Kit (Tiangen Technology Co., Ltd., Beijing, China). The extracted DNA was randomly fragmented, and paired-end libraries with an insert size of 350 bp were constructed. Sequencing was conducted on the Illumina HiSeq X platform with 150-bp pairedend reads, achieving an average coverage of five-fold, according to the manufacturer's guidelines. Additionally, resequencing data from two wild donkeys, *Equus africanus somaliensis* (Accession numbers: ERR650540-ERR650547, ERR650570-ERR650703) and *Equus kiang* (https://ngdc.cncb.ac.cn/biosample/browse/SAMC04905 2), were included [18].

Data processing

The donkey reference genome, GCA_016077325.1, was obtained from the NCBI website (https://www.ensem bl.org/) [19]. Quality control of the raw data was performed using Trimmomatic v.0.39 [20]. The clean data, consisting of 150-bp paired-end reads, were aligned to the GCA_016077325.1 reference genome using bwa v.0.7.17 [21, 22]. Multiple alignments were filtered with Samtools v.1.10 [23]. Single nucleotide polymorphism (SNP) calling, genotyping, variant detection, and the



Fig. 1 Geographic distribution of the studied Chinese donkey breeds showed in a partial map of China. HT and XJ indicate Hetian Gray donkeys and Xinjiang donkeys in the Xinjiang Uygur Autonomous Region (purple), respectively; YN, Yunnan donkeys in Yunnan province (pink); GZ, Guanzhong donkeys in Shaanxi province (light green); GL, Guangling donkeys in Shanxi province (deep green); DZ, Dezhou donkeys in Shandong province (blue); HBH, Huaibei Gray donkeys in Anhui province (yellow)

quality control were performed using HaplotypeCaller, SelectVariants, and VariantFiltration in GATK v.4.1.7.0 [24, 25]. The quality control criteria included filtering based on sequencing depth (QD < 2.0), mapping quality (MQ < 40.0), strand bias (FS > 60.0, SOR > 3.0), and rank sum tests (MQRankSum < -12.5, ReadPosRankSum < -8.0). These filtering criteria effectively eliminated lowquality variant sites. SNPs were subsequently filtered using plink v.1.9 [26] with the following criteria: (1) SNPs detection rates below 85%; (2) SNPs with Hardy-Weinberg equilibrium p-values $< 1 \times 10^{-7}$; (3) SNPs missing>10% of their genotype data; and (4) SNPs with a minor allele frequency < 5%. Additionally, individuals with >10% missing genotype data were excluded. After quality control, 1,435,541 SNPs remained for further analysis.

Population divergence

Based on the filtered SNPs, the merged SNP dataset was converted using vcftools v.0.1.16 [27]. When constructing the Neighbor-Joining (NJ) tree, we first extracted SNP sites using whole-genome resequencing data and calculated the genetic distance between these sites to construct a distance matrix. This matrix was then clustered using the NJ method, resulting in the NJ tree that reflects the genetic relationships among various donkey breeds. The NJ tree allows for an intuitive observation of the genetic relationships and distances between Hetian Gray donkeys and other donkey breeds. The NJ tree was constructed using the seqboot, dnadist, neighbor, and consense procedures in phylip v.3.698 [28]. The resulting tree file was uploaded to iTol for visualization [29].

A genetic relationship matrix was generated with gcta64 v.1.92.4 [30]. The principal component analysis (PCA) was conducted to compute the principal components of the genetic relationship matrix, and the results were visualized using the ggplot2 package in R v.4.4.0 [31].

ADMIXTURE analysis is a population structure analysis tool based on Bayesian statistical methods, which utilizes whole-genome resequencing data to infer the genetic composition and ancestral origins of each sample. In this study, we employed ADMIXTURE analysis to investigate the population structure of Hetian Gray donkeys alongside other donkey breeds, thereby elucidating the genetic composition and origins of the Hetian Gray donkeys. Ancestry clustering and genetic structure assessment were performed using ADMIXTURE v.1.3.0 [32], with the number of putative ancestors (K) ranging from 2 to 4. The optimal K value was estimated using cross-validation (CV), and visualization was conducted using the Structure Selector tool [33].

LD decay

PopLDdecay v.3.41 was employed to assess the linkage disequilibrium (LD) in the Hetian Gray donkey and other indigenous breeds [34]. The LD decay map was created using the Plot_MultiPop.pl script, which facilitated the analysis of selection intensity across subpopulations during evolution. The results were visualized using R scripts.

Identity by descent

IBDLD v.3.37 [35] was used to conduct an identity by descent (IBD) analysis. Haplotype frequencies shared between the Hetian Gray donkey and other donkey breeds within a 10,000 bp region were estimated using the software with the following parameters: "-plink bf int evolution-method GIBDLD - ploci 10 - nthreads 24 - step 0 - hiddenstates 3 - segment-length 10 - min 0.8." The normalized IBD (nIBD) was calculated as nIBD = cIBD/ tIBD, where cIBD represents the count of IBD haplo-types between the Hetian Gray donkey and other breeds, and tIBD is the total number of pairwise comparisons between them.

Estimation of possible ancestry with formal test of ADMIXTURE

The f4 ratio estimation calculates relatedness coefficients based on allele frequencies, providing insights into genetic differences and gene flow between populations. The f4 ratio algorithm (f4 ratio = f4(A, O; X, C)/f4(A, O; B, C)) from the ADMIXTOOLS Software Package was used to infer the ancestry of the Hetian Gray donkey [36]. In this analysis, the Hetian Gray donkeys were designated as the target population (X), with Xinjiang donkeys (C) as the presumed ancestor.

Selection signature in Hetian Gray donkeys

To identify genomic regions associated with selection in Hetian Gray donkeys, we calculated the population differentiation statistic (F_{ST}). F_{ST} values between Hetian Gray donkeys and Xinjiang donkeys were computed using a sliding 50-kb window with a 10-kb step in vcftools-0.1.16 [27]. To focus on the most significant region of the selection, genes located in the top 0.1% of regions with significantly high F_{ST} values were subsequently annotated to the donkey reference genome.

Results

Population genetic relationships between the Hetian Gray Donkey and other Chinese Indigenous Donkey breeds

The results of the NJ tree analysis (Fig. 2) demonstrated a clear clustering distinction between the CDBs and the wild donkeys. The phylogenetic tree revealed that the Hetian Gray donkeys clustered closely with both the Xinjiang donkeys and the Guanzhong donkeys, indicating close genetic links among these breeds. Consistently, the PCA results (Fig. 3) also showed that the Hetian Gray donkeys are clearly distinct from the outgroup, while clustering with other CDBs. Similar patterns were also observed in the ADMIXTURE analysis. The ADMIX-TURE ancestral component analysis revealed that the cross-validation (CV) error was minimized at K=2 (Fig. 4A), and the result indicates that the Hetian Gray donkeys share the similar genetic background with most of other CDBs including the Xinjiang donkeys and the Guanzhong donkeys. Nevertheless, to explore potential admixed population structures and avoid bias, results for K values as 3 and 4 were also presented (Fig. 4B). For instance, when K = 3, the Hetian gray donkey shows significant different components of ancestry from other studied populations, which may reflect that the breed had undergone isolation in the past, and the genetic similarity between the Hetian gray donkeys and the Xinjiang donkeys may reflect historical gene flow or shared ancestry at K = 4. These findings align with the results from the NJ tree and PCA, demonstrating that Hetian Gray donkeys and Xinjiang donkeys share a very similar genetic background, while a clear genetic distinction exists between the CDBs and wild donkeys. The analysis of LD revealed that the Hetian Gray donkeys have a more rapid LD decay than other breeds except for the Dezhou donkeys and the Yunnan donkeys (Fig. 5).

Identity by descent analysis of the Hetian Gray Donkey

In this study, IBD was used to explore the relationship between the Hetian Gray donkeys and other CDBs. The results revealed that the Hetian Gray donkeys showed highest shared IBD with Guanzhong donkeys, followed by Huaibei Gray donkeys, Xinjiang donkeys, Guangling donkeys, Dezhou donkeys and finally Yunnan donkeys (Fig. 6).

Estimation of possible ancestry with formal test of ADMIXTURE

The f4 ratio estimation algorithm was applied to estimate of possible ancestry of the Hetian Gray donkeys. In the output of the analysis, the greater the f4 ratio, the more likely it is that the B breed in the algorithm and the Xinjiang donkeys are the ancestors of the Hetian Gray donkeys. The results showed that the Xinjiang donkey and



Fig. 2 NJ tree of the Hetian Gray donkey and other studied donkey breeds. YN indicates Yunnan donkeys; HT, Hetian Gray donkeys; DZ, Dezhou donkeys; GL, Guangling donkeys; GZ, Guanzhong donkeys; XJ, Xinjiang donkeys; HBH, Huaibei Gray donkeys; kiang, Equus kiang; somali, Equus africanus somaliensis

the Guanzhong donkey are the most likely ancestors of the Hetian Gray donkey (Table 1).

Detection for signatures of selection

 F_{ST} was calculated for each SNP between the Hetian Gray donkeys and the Xinjiang donkeys. Annotation was carried out for the genes in the top 0.1% of the F_{ST} , and a total of 81 genes were identified, including the previously reported genes associated with body size, such as *TBX3*, *LCORL*, *NCAPC*, and *FAM184B* [37–39]. The strong selection was detected on EAA8 (Equus asinus autosome 8), EAA9 and EAA28. The most significant selection signature was detected in the region between 4,900,001 and 4,950,000 bp on EAA9, which harbors the conserved oligomeric Golgi complex subunit 6 (COG6) gene, a gene related with congenital disorders of glycosylation (CDG) in human beings (Fig. 7) [40, 41].

Discussion

The genetic relationships between the Hetian Gray Donkeys and other Donkey populations

The Hetian Gray donkeys are primarily distributed in Hetian County in the south of Xinjiang Uygur Autonomous Region. As Hetian was a prominent pivot along the Silk Road and donkeys played an important role in ancient commerce, it is crucial to investigate the genetic relationships between the Hetian Gray donkeys and other donkey breeds along this ancient route [42]. Due to the lack of systematic records documenting the history of the Hetian Gray donkey, it is a practicable resolution to reveal the origin of the Hetian Gray donkey with genomic data.

NJ tree, PCA, ADMIXTURE, and IBD analyses consistently indicate that the Xinjiang donkey and the Guanzhong donkey are the two indigenous donkey breeds most closely related to the Hetian Gray donkey. Most



Fig. 3 PCA for the studied donkey breeds. A and B showed the results of 3D and 2D PCA, respectively. YN indicates Yunnan donkeys; HT, Hetian Gray donkeys; DZ, Dezhou donkeys; GL, Guangling donkeys; GZ, Guanzhong donkeys; XJ, Xinjiang donkeys; HBH, Huaibei Gray donkeys; kiang, *Equus kiang*; somali, *Equus africanus somaliensis*



Fig. 4 ADMIXTURE analysis for the studied donkey breeds. (A) Cross-validation for each run in ADMIXTURE analysis. (B) ADMIXTURE analysis results of the studied donkey breeds. YN indicates Yunnan donkeys; HT, Hetian Gray donkeys; DZ, Dezhou donkeys; GL, Guangling donkeys; GZ, Guanzhong donkeys; XJ, Xinjiang donkeys; HBH, Huaibei Gray donkeys; WD, Wild donkeys

local donkey breeds tended to cluster together in PCA, which is consistent with the results of the ADMIXTURE. These results may be attributed to common ancestral lineages shared by most CDBs. Historical records suggest that earliest domesticated donkeys in China were found in southern Xinjiang, and then spread into other regions through Gansu and Shaanxi provinces [43–45].

As the Hetian Gray donkey is the only large-body-size donkey breed in the south of Xinjiang, a hypothesis was proposed that the Hetian Gray donkeys arose from the hybridization of wild donkeys with Xinjiang donkeys [2]. In the present study, we included *Equus kiang* as its distribution overlaps that of the Hetian Gray donkey and the Xinjiang donkey. However, our results showed that the Hetian Gray donkeys have distant genetic relationship with the *Equus kiang*, which is the only wild donkey population distributed in the south of Xinjiang province, and thus denied the hypothesis. Based on our genomic evidence, the Hetian Gray donkeys have close genetic relationship with domestic donkey breeds, especially the Xinjiang donkey and the Guanzhong donkey, and our results suggest that the Hetian Gray donkeys have a native origin.

The impact of the ancient trade routes on the Hetian Gray donkeys

The distribution area of the Hetian Gray donkey breed is located at the core of the ancient Silk Road, a trade route that spanned from China to the Eurasian continent [46]. The Silk Road was not only a channel for the exchange of goods but also a crucial route for the migration of people and livestock. Donkeys, as important pack animals at the time, were widely used for transportation along the route.





Fig. 5 LD decay of the studied donkey breeds. The studied populations are indicated with different colors. YN indicates Yunnan donkeys; HT, Hetian Gray donkeys; DZ, Dezhou donkeys; GL, Guangling donkeys; GZ, Guanzhong donkeys; XJ, Xinjiang donkeys; HBH, Huaibei Gray donkeys



Fig. 6 The estimation of IBD shared between the Hetian Gray donkey and other Chinese indigenous donkey breeds. YN, Yunnan donkeys; DZ, Dezhou donkeys; GL, Guangling donkeys; GZ, Guanzhong donkeys; XJ, Xinjiang donkeys; HBH, Huaibei Gray donkeys

Their endurance, adaptability, and ease of care made them indispensable pack animals of the trade.

The original distribution area of the Guanzhong donkey breed was near Xi'an, the eastern end of the Silk Road. Xi'an was not only a major capital during the Han and Tang dynasties but also a prosperous commercial center of the Silk Road, and there were close trade links between Xi'an and Hetian [47]. These frequent trade and migration activities between the two areas may promote the exchanges of livestock along the ancient trade route, and may have facilitated gene flow into the Hetian area, ultimately contributing to the formation of the Hetian Gray donkey (Fig. 8).

Though comprehensive analyses of genomic data were performed regarding the genetic relationship between the Hetian Gray donkeys and other donkey populations in the present study, some limitations should be addressed. The unequal sample size of the studied populations may lead to bias of the LD decay results. The sequencing depth of most studied populations is $5\times$, which limits some in-depth analysis of the genomic data. Further studies, using larger and even sample sizes and deep sequencing data, and ideally including archeological evidence, are still needed to reveal the details of the origin of the Hetian Gray donkeys.

Selective signature at the genome level of the Hetian Gray Donkey

Population analyses have revealed that the Hetian Gray donkey has the closest genetic relationship with the Xinjiang donkey. However, they are regarded as two distinctive breeds due to the much different phenotypes of exterior appearance between them. It was speculated that the two breeds had been selected differentially for the phenotypes, though they may origin from common ancestries [3]. Our analysis of the genomic selective signature showed that the genes related with body sizes had been subjected to selection for the two populations, and it is accordant with the fact that the Hetian Gray donkeys feature their larger body sizes while the Xinjiang donkeys usually have a small stature. Besides the known genes associated with the body sizes, such as TBX3, LCORL, NCAPC, and FAM184B [37-39], a new gene COG6 was identified in the most significant selected region of the

Table 1 Estimation of the genome-wide possible ancestry of the Hetian Gray donkeys

Α	0	Х	С	Α	0	В	с	F4 Rate	Std.err	Z(null=0)
GL	WD	HT	XJ	GL	WD	GZ	XJ	1.4017	0.0808	17.36
GL	WD	HT	XJ	GL	WD	YN	XJ	0.7876	0.0277	28.43
GL	WD	HT	XJ	GL	WD	DZ	XJ	0.6225	0.0170	36.52
GL	WD	HT	XJ	GL	WD	HBH	ΧJ	0.6893	0.0194	35.56

Note: A, the least likely ancestor of the Hetian Gray donkey, and the Guangling donkey (GL) was selected as A; O, wild donkey (WD); X, target breed, Hetian Gray donkeys (HT); C, the most likely ancestor of the Hetian Gray donkey, the Xinjiang donkey (XJ); B, the possible ancestor to be tested, including the Guanzhong donkey (GZ), the Yunnan donkey (YN), the Dezhou donkey (DZ) and the Huaibei Gray donkey (HBH)



Fig. 7 Manhattan plot of the F_{ST} values between the Hetian Gray donkeys and the Xinjiang donkeys. The F_{ST} values are calculated for each 50 kb autosomal window. Adjacent chromosomes were indicated with different colors to make a better distinction of the borders



Fig. 8 The migration of the Guanzhong donkey along the Silk Road trade routes showed in a partial map of China. The red dot indicates Xi'an city; the orange dot, Hetian area; the white line, the Silk Trade Routes within China; The red dashed line, the possible migration route of the Guanzhong Donkey; Shaanxi province, Gansu province and the Xinjiang Uygur Autonomous Region are represented in green, light green, and light blue, respectively

genome. COG6-CDG is a kind of disorder caused by *COG6* deficiency in human beings. COG6-CDG leads to retardation, developmental disability, and other symptoms [48], which suggests the function of *COG6* related with body growth. *COG6* may serve as a new candidate gene for body sizes in donkeys.

The conservation of the Hetian Gray Donkey

The Hetian Gray donkeys were officially identified as a new breed in 2006 [3]. It had been used as labor animals but now more and more Hetian Gray donkeys are bred for milk and meat production. In this study, the Hetian Gray donkeys exhibited a low decay rate of the linkage disequilibrium compared to most other CDBs. This suggests that the Hetian Gray donkey has not undergone intensive breeding activities, and has the potential to be further improved on its economic traits. However, there are now only approximately 3,000 Hetian Gray donkeys in Hetian area, and the population is still decreasing in recent years. Efficient conservation efforts should be made to protect this special local breed.

Conclusion

The Hetian Gray donkeys share a close genetic relationship with indigenous Chinese donkey breeds, and Xinjiang donkeys and Guanzhong donkeys played a significant role in the formation of the Hetian Gray breed. The genes related with body sizes were selected in the breeding activity of the Hetian Gray donkeys, and contributed on the large stature of the breed.

Abbreviations

Chinese indigenous donkey breed
Chinese indigenous donkey breeds
Single nucleotide polymorphisms
Single nucleotide polymorphism
Linkage disequilibrium
Cross-validation
Principal Component Analysis
Identity by descent
Neighbor-Joining
Equus asinus autosome 8
Component of Oligomeric Golgi Complex 6
Congenital disorders of glycosylation

Acknowledgements

The authors thank the assistances of Xiaoling Zhou from Tarim University and Fang Tian from Kunlunlvyuang Co. for collecting the samples of the Hetian Gray donkeys, and appreciate the support from Xiaoran Lin, YugangHe, and Ying Li for the data analysis. This research was supported by High-performance Computing Platform of China Agricultural University.

Author contributions

YYL: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Visualization, Writing– original draft; YLiu: Methodology, Formal analysis, Investigation, Writing– original draft; review & editing. SQL: Methodology, Formal analysis, Investigation, Writing– original draft; review & editing. YGH: Methodology, Investigation, Writing– review & editing. XRL: Methodolog– Writing– review & editing. YL: Resources, Writing– review & editing. TY: Resources, Writing– review & editing. TY: Resources, Writing– review & editing. MF: Resources, Writing– review & editing. HTZ: Resources, Writing– review & editing. MHZ: Resources, Writing– review & editing. MF: Resources, Writing– review & editing. CJZ: Conceptualization, Methodology, Supervision, Funding acquisition, Writing– review & editing.

Funding

This study was financially supported by the Program for Changjiang Scholars and Innovative Research Team in University (Grant No. IRT1191), the Project on the Third National Survey of the Livestock and Poultry Genetic Resources (Grant No. 19221073), the Public Science and Technology Research Funds Projects of Agriculture (Grant No. 201003075), the Beijing Key Laboratory for Genetic Improvement of Livestock and Poultry (Grant No. 2171100002217072), and the project of the National Germplasm Center of Domestic Animal Resources. Donkey Innovation Team of Shandong Modern Agricultural Industry Technology System, grant number SDAIT-27.

Data availability

The genome resequencing data for the Xinjiang donkey are available upon request. In addition, the whole-genome resequencing data for the Guangling donkey and Dezhou donkey populations have been deposited in GenBank under BioProject accession PRJNA800192. Furthermore, the whole-genome

resequencing data for the Yunnan donkey, Huaibei Gray donkey, Guangling donkey, and Hetian Gray donkey populations are available under BioProject accession PRJNA1215270.

Declarations

Ethics approval and consent to participate

The samples were obtained following the principle approved by the Animal Care and Use Committee of China Agricultural University (permit number: XK257). The informed consent from the owners were obtained for using the samples of the donkeys in the present study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 January 2025 / Accepted: 10 April 2025 Published online: 30 April 2025

References

- Zhou Z, Fan Y, Wang G, Lai Z, Gao Y, Wu F, Lei C et al. Detection of selection signatures underlying production and adaptive traits based on Whole-Genome sequencing of six Donkey populations. Animals (Basel). 2020;10(10):1823.
- Mamutiang K, Minawal, Li X. Performance analysis and discussion of the Xinjiang Hetian Gray Donkey. Xinjiang Anim Husb. 2012;03:37–8.
- 3. China National Commission of Animal Genetic Resources. Animal genetic resources in China horses donkeys Camels[M]. China Agriculture; 2011.
- Wei H, Yang F, Li P. A brief discussion on the current status and strategies for the conservation of Hetian green Donkey genetic resources. Xinjiang Anim Husb J. 2011;01:53–4.
- Bhati M, Kadri NK, Crysnanto D, Pausch H. Assessing genomic diversity and signatures of selection in original Braunvieh cattle using whole-genome sequencing data. BMC Genomics. 2020;27:1–14.
- Ghildiyal K, Nayak SS, Rajawat D, Sharma A, Chhotaray S, Bhushan B, et al. Genomic insights into the conservation of wild and domestic animal diversity: a review. Gene. 2023;886:147719.
- Xia X, Zhang S, Zhang H, Zhang Z, Chen N, Li Z, et al. Assessing genomic diversity and signatures of selection in Jiaxian red cattle using wholegenome sequencing data. BMC Genomics. 2021;22:43.
- Zhang S, Yao Z, Li X, Zhang Z, Liu X, Yang P, et al. Assessing genomic diversity and signatures of selection in Pinan cattle using whole-genome sequencing data. BMC Genomics. 2022;460:1–10.
- Wang G, Wang F, Pei H, Li M, Bai F, Lei C, Dang R. Genome-wide analysis reveals selection signatures for body size and drought adaptation in Liangzhou Donkey. Genomics. 2022;6:110476.
- 10. Xiong X, Liu J, Rao Y. Whole genome resequencing helps study important traits in chickens. Genes (Basel). 2023. 14(6).
- Kumar A, et al. Whole genome resequencing revealed genomic variants and functional pathways related to adaptation in Indian Yak populations. Anim Biotechnol. 2024;35(1):2282723.
- Chai ZX, et al. Whole-genome resequencing provides insights into the evolution and divergence of the native domestic Yaks of the Qinghai-Tibet plateau. BMC Evol Biol. 2020;20(1):137.
- Shen J, Yu J, Dai X, Li M, Wang G, Chen N, Chen H, Lei C, Dang R. Genomic analyses reveal distinct genetic architectures and selective pressures in Chinese donkeys. J Genet Genomics. 2021;48:737–45.
- Chen J, et al. Single nucleotide polymorphisms (SNPs) and indels identified from whole-genome re-sequencing of four Chinese Donkey breeds. Anim Biotechnol. 2023;34(5):1828–39.
- Zhou Z et al. Detection of selection signatures underlying production and adaptive traits based on Whole-Genome sequencing of six Donkey populations. Anim (Basel), 2020. 10(10).
- Liu Y, et al. Genetic architectures and selection signatures of body height in Chinese Indigenous donkeys revealed by next-generation sequencing. Anim Genet. 2022;53(4):487–97.

- 18. Schoch CL, et al. NCBI taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). 2020;2020:baaa062.
- Wang C, Li H, Guo Y, Huang J, Sun Y, Min J, et al. Author correction: Donkey genomes provide new insights into domestication and selection for coat color. Nat Communication. 2021;1:725.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30:2114–20.
- 21. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25:1754–60.
- 22. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. Bioinformatics. 2010;26:589–95.
- 23. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of samtools and BCFtools. GigaScience. 2021;10:giab008.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing nextgeneration DNA sequencing data. Genome Res. 2010;20:1297–303.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protocols Bioinf. 2013;43:1101–11033.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
- 27. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. 1000 Genomes project analysis group. The variant call format and vcftools. Bioinformatics. 2011;27:2156–8.
- 28. Felsenstein J. 1996.PHYLIP. https://evolution.genetics.washington.edu/phylip. html
- Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.
- 30. Yang J, Lee SH, Goddard ME, Visscher. P.M.GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.
- Wickham H ggplot2: elegant graphics for data analysis. New York: Springer-Verlag., Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. American Journal of Human Genetics. 2016; 88: 76–82.
- 32. Alexander DH, Novembre J, Lange K. Fast modelbased Estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
- Li YL, Liu JX. StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour. 2018;18:176–7.
- 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;35:1786–8.

- Han L, Abney M. Identity by descent Estimation with dense genome-wide genotype data. Genet Epidemiol. 2011;35:557–67.
- 36. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history.
- Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. Nature. 2022;7933:704–12.
- Akiyama M, Ishigaki K, Sakaue S, Momozawa Y, Horikoshi M, Hirata M, et al. Characterizing rare and low-frequency height-associated variants in the Japanese population. Nat Communication. 2019;1:4393.
- Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. Am J Hum Genet. 2019;1:65–75.
- Mandel H, COG6-CDG, et al. Expanding the phenotype with emphasis on glycosylation defects involved in the causation of male disorders of sex development. Clin Genet. 2020;98(4):402–7.
- Xia ZJ, et al. The Swedish COG6-CDG experience and a comprehensive literature review. JIMD Rep. 2023;64(1):79–89.
- 42. Taylor W, et al. Early pastoral economies along the ancient silk road: biomolecular evidence from the Alay Valley, Kyrgyzstan. PLoS ONE. 2018;13(10):e0205646.
- China National Commission of Animal Genetic Resources. Animal genetic resources in China horses donkeys Camels[M]. p18-19. China Agriculture; 2011.
- 44. Qinglan Ge. Maternal origin and genetic diversity of Chinese domestic donkeys. Thesis for Masters Degree, Northwest A & F University. 2007.
- Liu YF, Dong NN. Review and reflection on the domestication of donkeys. South Cult Relics. 2024;03:201–20.
- Liu X, Shaffer LJ. Silk road: monuments of China's caravan route. New York: Berghahn Books; 2007.
- Zhang Q, Yuan J. The genetic impact of ancient trade routes on the formation of domestic animal populations in central Asia. Public Libr Sci ONE. 2015;3:e0123768.
- Lugli L, Bariola MC, Ferri L, Lucaccioni L, Bertucci E, Cattini U, et al. Disorder of sex development associated with a novel homozygous nonsense mutation in COG6 expands the phenotypic spectrum of COG6-CDG. Am J Med Genet A. 2021;185(4):1187–94.

Publisher's note

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