

Genomic Analyses Reveal Distinct Genetic Architectures and Selective Pressures in Chinese Donkeys

Jiafei Shen

Northwest A&F University: Northwest Agriculture and Forestry University

Jie Yu

Northwest A&F University: Northwest Agriculture and Forestry University

Xuelel Dai

Northwest A&F University: Northwest Agriculture and Forestry University

Gang Wang

Northwest A&F University: Northwest Agriculture and Forestry University

Ningbo Chen

Northwest A&F University: Northwest Agriculture and Forestry University

Chuzhao Lei

Northwest A&F University: Northwest Agriculture and Forestry University

RUIHUA DANG (✉ drh211@hotmail.com)

Northwest A&F University: Northwest Agriculture and Forestry University <https://orcid.org/0000-0002-4818-5871>

Research

Keywords: Donkey, whole-genome resequencing, NCAPG, LCORL

Posted Date: November 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-111083/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Donkey (*Equus asinus*) is an important livestock animal in China because of its nourishment and medical value. After a long period of natural and artificial selection, the variety diversity and phenotype of donkey are very rich. However, due to the lack of genome-wide studies, its genomic value remains unclear.

Results

We clarified the genetic and demographic characteristics of Chinese donkeys and the selective pressures on them by analyzing 78 whole genomes from 12 breeds. We found that Gunsha donkey had the lowest nucleotide diversity, longest length, and largest number of runs of homozygosity. Phylogenetic analysis revealed obvious geographical distribution trends in Chinese donkeys. In selective sweep, gene annotation, functional enrichment, and differential expression analyses between large and small body donkey groups, we identified selective signals, including *NCAPG* and *LCORL*, to be related to rapid growth and large body size.

Conclusion

Our findings improve the understanding of the evolutionary history and formation of different donkey breeds and provide theoretical insights into the genetic mechanism underlying breed characteristics and molecular breeding programs of donkey clades.

Background

The distribution of domestic donkey, an important farm and transportation animal in China, is concentrated between 32° and 42° north latitude. Domestic donkeys live in the northwest, north, and southwest regions of China in temperate and warm climates (1). After a long period of natural and artificial selection, domestic donkeys developed to have different body shapes and production performance. They are particularly suited for transport in mountainous and arid regions. With advancements in agricultural mechanization and rapid development of the transportation industry, domestic donkeys have gradually been eliminated as a means of transportation, and the number of donkey herds has decreased sharply; however, the meat and medicinal values of domestic donkeys have increased (2).

The history and origins of the donkey are particularly interesting. Because the earliest archeological evidence (fossil evidence) of domestic donkeys was found in Egypt, it was originally thought that domestic donkeys were domesticated from Nubian wild donkeys around the Nile in ancient Egypt (3–6). However, ancient donkey bone samples were found in Syria, Iran, and Iraq in the 1980s because of technical limitations at the time. Some archeological researchers even believed that the Asian wild donkeys were the direct ancestors of domestic donkeys (7–11). Donkeys can be either domestic or wild.

Ten articulated donkey skeletons were discovered in the three brick tombs, causing scholars to believe that domestic donkeys were domesticated from wild African donkeys (12). Archeological, historical, and ethnographic data indicated that there were at least three different groups of wild asses in Africa 2000 years ago, with only two surviving until modern times (12). Somali wild ass is critically endangered but is still found in Somalia, Ethiopia, and Eritrea. Nubian wild asses have rarely been seen in recent years, creating concerns that these populations have become extinct. Previous studies of mitochondrial DNA levels have shown that modern donkeys can be clearly separated into two distinct clades: clade I was clustered with the Nubian wild ass (*Equus africanus africanus*) and clade II was closer to the extinct Somali wild ass (*E. africanus somaliensis*) (13, 14).

Whole-genome sequencing is a powerful method for evaluating population structures and demographic patterns as well as for identifying regions associated with important economic and environmental adaptive traits. However, because of the lack of a complete chromosome-level reference genome, most recent studies of donkeys have been based on mitochondrial levels. Considering that the entire genome variation of donkey is largely unexplored, we sequenced the whole genome of 78 donkeys from 12 breeds of different geographic origins to examine the genomic diversity, population structure, and demographic history of this important livestock species and to reveal possible signs of natural and artificial selection.

Methods

Ethics statement

This study was conducted to the guidelines of the Council of China and animal welfare requirements. Based on the recommendations of the Regulations for the Administration of Affairs Concerning Experimental Animals of China, the Institutional Animal Care and Use Committee of Northwest A&F University approved all animal experiments.

Sample collection and sequencing

We evaluated 78 donkeys from different locations (Table S1). Genomic DNA was extracted from the ear tissue or blood samples using the standard phenol-chloroform protocol (15), amplified in genomic libraries with an average insert size of 500 base pairs, and sequenced (150-base pair paired-end reads) on an Illumina HiSeq 2000. Additional details are provided in Table S2

Pseudo chromosomal level reference genome assembly

We downloaded the latest donkey reference genome (PRJNA259598) from NCBI. This genome was published in 2015 and is currently at the scaffold level (16). From an evolutionary perspective, horses are relatives of donkeys and high-quality reference genomes of horses have been determined at the chromosomal level. Therefore, we used the reference genome of a related horse species (PRJNA421018) to assemble the donkey scaffold at the chromosome level (17). First, we used minimap2 software (18) (parameter: secondary = no -cx asm10) to compare the donkey's reference genome to the horse's reference genome to obtain a paf file. Next, using a Python script (Connect_Pseudo_Chromosome.py,

Note S1), according to the comparison information in the paf file, the donkey's Scaffold was assembled to at the pseudo chromosome level, and the connection information of each scaffold was recorded. To obtain annotation information for the donkey's pseudo genome, we used a Python script (Pseudo_Chromosome_Annotation.py, Note S2) to convert the annotation information at the donkey scaffold level to the pseudogene which was assembled based on the donkey's scaffold connection information.

Alignments and variant identification

All cleaned reads were aligned to the reference genome (PRJNA421018) linked to “30 + X + Y + unplaced” pseudo-chromosomes (Table S3) using BWA-MEM (19) with default settings. Duplicate reads were filtered using Picard tools. SNPs were detected with the Genome Analysis Toolkit (GATK, version 3.6-0-g89b7209) (20) and filtered using the “VariantFiltration” tool, as described in the Supplementary Information.

Phylogenetic and population structure analyses

The NJ tree, PCA, and ADMIXTURE methods were used to explore the genetic relationships among key populations. An individual-based NJ tree based on the matrix of pairwise genetic distances from the autosomal SNP data of 60 donkeys was constructed with PLINK (version 1.9) and visualized in FigTree. The TreeMix program was used to evaluate the population-level phylogeny (21). PCA was performed using the SmartPCA program in the EIGENSOFT v5.0 package (22) and the significance of the eigenvectors was detected by the Tracy-Widom test. The population genetic structure was estimated using ADMIXTURE v. 1.3.0 (23) considering 2–10 clusters (K).

Genome-wide selective sweep test

To detect selective sweeps in different body size groups, we compared two types of donkeys: (i) large body size group as the reference and small body size group as the object population; (ii) small body size group as the reference and large body size group as the object population. Two methods were used: (i) fixation index (F_{ST}) values (24) were calculated in sliding 50-kb windows with 20-kb steps along the autosomes using VCFtools (25); (ii) high differences in genetic diversity (π ln ratio) were calculated with 50-kb sliding-windows and 20-kb steps along the autosomes using VCFtools and in-house scripts. Significant genomic regions were identified as having a P -value < 0.005 . The two methods showed outlier signals (P -value < 0.005) in overlapping regions and were therefore considered as candidate selective regions. KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/anno_iden.php) was used to examine their biological functions and pathways involved in these regions.

Results

Analyses of 78 China donkey genomes revealed > 24.4 million single-nucleotide polymorphisms (SNPs)

Seventy-eight donkey genomes were obtained, representing 34 East China plain donkeys (Guangzhong, Taihang, Dezhou, Huaibehui, Biyang), 25 loess plateau donkey (Qinghai, Gunsha, Jiami, and Qingyang), and 19 Southwest China plateau donkey (Guoluo, Xinjiang, and Xizang) (Fig. 1A), with $\sim 12.56 \times$ sequence coverage for each individual. Comparisons among the genome sequences revealed 24.4 million putative autosomal SNPs, with 67.39% mapping to intergenic regions, 31.14% to intronic regions, and only 0.96% mapping to exonic regions of the genome. Functional annotation of the SNPs assigned to protein-coding regions identified 50.75% SNPs that produced nonsynonymous amino acid substitutions and 46.62% SNPs causing synonymous substitutions, with 1.1% SNPs resulting in the gain or loss of a stop codon (Fig. 1A). Analysis of the genomic characteristics showed that nucleotide diversity was highest in Huaibehui donkey and lowest was in Gunsha donkey, while diversity was similar in other donkey breeds (Fig. 1C). Runs of homozygosity (ROH) analysis performed using the sum of the ROH (S_{ROH}) and number of ROH (N_{ROH}) showed significant differences in 12 donkey breeds. The Gunsha donkey was characterized as having the highest S_{ROH} (28,324.9–4835.47 Kb) and highest N_{ROH} . The Qinghai donkey showed the lowest S_{ROH} of 6172.23–9004.23 Kb.

Population genetic structure and relationship

Neighbor-joining (NJ) trees, principal component analysis (PCA), and ADMIXTURE were used to explore the genetic relationships among the examined donkey populations. In the NJ tree, the donkey breeds were separated into three main clades: East China Plain donkey, loess plateau donkey, and plateau area donkey (Fig. 2A). PCA showed that the first PC explained 2.66% of the total variation and separated Gunsha and Jiami donkeys from other donkey breeds. The second PC, explaining 2.59% of the total variation, could not separate any breed (Fig. 2B). Admixture analysis did not distinguish between each donkey breed and different ancestral types. The results showed that most individuals were of mixed ancestry (Figure S1).

Two clearly separated Chinese donkey mitochondria clades

Previous studies divided the phylogenetic relationship of donkey mitochondria in two lineages. Thus, we constructed a highly supported phylogenetic tree based on the 78 mtDNA genome sequences (coverage $> 100\times$) from this study and 8 mtDNA genome sequences downloaded from NCBI (including one horse, four *E. hemionus*, one *E. kiang*, and two *E. asinuss somali*).

As outgroup, horses were independent of domestic and wild donkeys. Similarly, Asian wild asses (*E. hemionus* and *E. kiang*) formed a branch between the horse and domestic donkey. Two highly divergent phylogenetic clades (Clade I and Clade II) were identified in domestic donkeys, with two Somali wild asses clustering within a group close to Clade II. In addition to mtDNA genome phylogeny analysis, we constructed a reduced median network based on 78 domestic genome sequences to assess whether their evaluation patterns were similar between the two lineages. Seventy-eight samples were defined as having 12 haplotypes based on their variety information. Similar to mitochondrial cluster analysis, the haplotypes can be divided into two branches; Clade I showed a simple star-like shape, whereas the

genetic architecture of Clade II haplotypes was more complex, with more universally occurring haplotypes.

Genome-wide differential selection in small and large body size

The phenotypes of Chinese domesticate donkey we found to be very rich and were divided into three groups according to body size (large donkeys, medium donkeys, and small donkeys). We applied the F_{st} and π -ratio to evaluate genomic regions related to the selection of the large and small body size groups. The Manhattan plots showed the distribution trend of the selected signals among these different groups (Fig. 3A). The 1% top regions from different methods were considered as candidate selection regions, and 332 and 589 genes were obtained after annotation, respectively. Only 85 overlapping genes (Fig. 3D) identified by the two selected statistical methods were defined as candidate genes under positive selection (Table S4). In both methods, the most significant signals were observed on chromosome 3, which mainly contains two genes, non-SMC condensin I complex subunit G (*NCAPG*) and ligand-dependent nuclear receptor corepressor-like (*LCORL*). Previous studies reported that *LCORL* and *NCAPG* are closely related to body size traits in other species (26–29). We also found a well-known gene, *FAM184B*, which is related to the body size traits on the 3 chromosomes from the π -ratio (30, 31). In addition, 11 and 5 nonsynonymous sites were found on *LCORL* and *NCAPG*, respectively, through mutation site analysis. According to genotype analysis of 16 nonsynonymous loci, the nonsynonymous genotypes significantly differed between large and small donkeys (Fig. 3B). *LCORL* and *NCAPG* were observed as neighboring genes at the chromosome level, occupied near 0.2-Mb regions, and contained 586 SNPs overall. We calculated the linkage disequilibrium values of these SNPs to determine whether the SNPs were linked, which revealed that SNPs in this region have strong linkage (Fig. 3C).

In addition, we calculated the SweepD values of 9 representative donkey breeds to choose the significant selection signals of each breed. Experience P -values of SweepD value less than 0.01 were considered as positive selection regions for each breed, and the distribution of SNPs in each breed was displayed in Manhattan plots (Figure S2). By annotating the candidate regions of each variety, the counts of candidate genes for each breed were displayed on a histogram, as shown in Fig. 3E. Intersections of candidate genes in each breed were considered, and the *NCAPG* and *PTPRN2*, which are related to body size traits, were detected in large donkeys (Dezhou, Guanzhong, and Guoluo donkey). Finally, we performed Kyoto Encyclopedia of Genes and Genomes and Gene Ontology enrichment analysis of the 85 candidate genes to determine the known functional information for each gene (Figure S3).

Discussion

In this study, we analyzed the whole genome sequence of 78 donkeys from 12 breeds. According to their body sizes, 12 donkey breeds were divided into groups of small donkeys (Xinjiang, Tibetan, Qinghai, Haibehui, Taihang, and Gunsha donkeys), medium donkeys (Jiami and Qingyang donkeys), and large donkeys (Biyang, Dezhou, Guanzhong, and Guoluo donkeys). The genetic diversity of these animals may

reflect various evolutionary events, including bottlenecks during domestication, introgression, recent cross-breeding, and artificial selection. Gunsha donkey showed reduced levels of nucleotide diversity compared to the other breeds, which is likely the result of intensive artificial selection in China. The nucleotide diversity of the other 11 donkey breeds was similar, which may be explained by unknown historical demography such as population expansion or introgression and a degree of inbreeding or a smaller effective population size. The length and number of homozygosity (ROH) among different donkey breeds were analyzed. ROH was first described by Gibson (32) and defined as contiguous homozygous genotype segments in the genome that are present in an individual because of the transmission of identical haplotypes from parents to their offspring (33–35). Inbreeding can increase the homozygosity of a population; as the level of inbreeding increases, the possibility for becoming homozygous for harmful recessive genes also increases, which may lead to decreases in reproduction, viability, and phenotypes of the offspring (36). The segment and number of ROH in Gunsha donkey were longer and larger, respectively, than those in other donkey breeds, possibly because of strong artificial selection and recent excessive inbreeding of Gunsha donkeys. Qinghai donkey had the shortest and smallest number of ROHs among the 12 donkey breeds. This may be because of the following reasons: 1. Qinghai donkey inbreeding may have been low in the large area of Qinghai and 2. Qinghai donkey was not the only or most important economic and transportation animal in the local area, and thus the intensity of natural and artificial selection may be relatively low. A lower inbreeding coefficient and lower selection pressure may be underlying factors causing Qinghai donkey to have the smallest number of and shortest ROH.

Population structure and phylogenetic analyses can improve the understanding of the evolutionary process of certain species or breeds. As there is no complete chromosome-level reference genome available, no population genomic analyses have examined the genetic ancestry and population structure of Chinese donkey. In our study, we selected representative breeds of Chinese donkey to analyze their relationships. According to the NJ tree, Mongolian wild ass was the outer group, and Chinese donkey breeds were clustered into three branches according to their geographic region. These mainly included the East China Plain, loess plateau, and plateau area donkeys. The results of PCA showed that only the first PC separated Gunsha and Jiami donkey from the other donkey breeds, whereas other PCs could not separate the breeds. Admixture analysis did not distinguish between each donkey breed and different ancestral types. The reason that clear results were not obtained by PCA and admixture analysis may be that most individuals were of mixed ancestry. With the rapid development of civilization, donkeys were rarely used as a means of transport; however, domestic donkeys have been widely used for their meat and medicinal value. This has led to the artificial migration, hybridization, and breeding of Chinese donkeys. Such mixed results can be explained by mixed ancestry between donkey breeds. The degree of genetic diversity among Chinese donkey breeds may not be high, and the difference in genetic background is not obvious. Similar to the results of previous studies, we found that the Chinese donkey has two maternal sources and forms two branches, clade I and clade II. Among them, Clade II and Somali wild donkeys gather together, and this branch may have originated from Somali wild donkeys in Africa.

As the Asian wild ass and Chinese donkey were present on different branches, it does not appear that the domestic donkey originated from the Asian wild ass.

By performing whole-genome selection scans, we analyzed the genetic basis of the divergence of different body sizes of donkey groups. The top two divergent genes on chromosome 3 between the large and small donkey groups were both strongly related to cell mitosis and bone development, including *NCAPG* and *LCORL*, which may play an important role in body growth. *NCAPG* is a subunit of the condensin 1 complex which is involved in chromosome condensation and interacts with a DNA methyltransferase linking methylation and chromatin condensation (37). *LCORL* encodes a transcription factor that is thought to function during spermatogenesis (38). These neighboring genes exhibit strong linkage. In previous studies of the functions of these two genes, the regions containing the genes were often described together in many species. The regions of *NCAPG* and *LCORL* have been identified as loci for adult human height in European (39–43), Japanese (44) and African/African American populations (45, 46). The region was also associated with the peak height velocity in infancy (47), birth weight (48) and birth length (49). In cattle, the two gene regions were associated with birth weight (50), weaning and yearling weight (51), increase in body frame size (52) and carcass weight (53). Further, this region was found to be associated with height in some horse breeds (26, 29, 54, 55) and has been identified as highly differentiated between dog breeds (38) and a selective sweep region in European domestic pigs (56).

Conclusions

In our study, whether it is the selection signal among different body types of donkeys, or the selection signal in large-sized donkey breeds. *NCAPG* and *LCORL* were in significant selection regions. Thus, in the process of natural selection and artificial domestication of donkeys in China, these two genes may play a vital biological role in the evolution of donkeys.

Abbreviations

BY: Biyang donkey

DZ: Dezhou donkey

GZ: Guanzhong donkey

TH: Taihang donkey

HL: Huaibehui donkey

QH: Qinghai donkey

GS: Gunsha donkey

JM: Jiami donkey

QY: Qingyang donkey

GL: Guoluo donkey

XJ: Xinjiang donkey

XZ: Xizang donkey

Declarations

Acknowledgement

We thank for Editage (www.editage.com) for English language editing.

Funding

This work was supported by the Key Research & Development Plan of Shaanxi Province of China (2020NY-002), Dong-E-E-Jiao Co. Ltd (20191001), and National Natural Science Foundation of China (81770514).

Availability of data and materials

The raw whole genome sequencing data has been available at the NCBI Short Read Archive under the BioProject accession number PRJNA675210.

Author information

Affiliations

Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China

Jiafei Shen, Jie Yu, Xuelei Dai, Gang Wang, Ningbo Chen, Chuzhao Lei, Ruihua Dang

Contributions

Ruihua Dang and Chuzhao Lei designed the study and supported the funding. Jiafei Shen curated and analyzed the data. Jiafei Shen wrote the original manuscript. Xuelei Dai reviewed the manuscript. Jie Yu and Gang Wang organized sampling and conducted fieldwork. All authors commented on the manuscript and gave final approval for publication.

Ethics approval and consent to participate

This study was approved by Institutional Animal Care and Use Committee of Northwest A&F University following the recommendation of the Regulations for the Administration of Affairs Concerning Experimental Animals of China.

Consent for publication

Not applicable.

Competing interests

The authors declare no potential conflict of interest for this study.

References

1. Chang H. Animal Genetic Resource in China: Horses, Donkeys and Camels: China Agricultural Press; 2012.
2. Li Y, Fan Y, Shaikh AS, Wang Z, Wang D, Tan H. Dezhou donkey (*Equus asinus*) milk a potential treatment strategy for type 2 diabetes. *Journal of ethnopharmacology*. 2020;246:112221.
3. Boessneck J, von den Driesch A. Tell el-Dabā: Tiere und historische Umwelt im nordost-Delta im 2. Jahrtausend v. Chr. anhand der Knochenfunde der Ausgrabungen 1975-1986. VII: Verlag der Österreichischen Akademie der Wissenschaften; 1992.
4. Petrie WMF. Roman Portraits and Memphis IV, Tarkhan I and Memphis V, Tarkhan II: Cambridge University Press; 2013.
5. Boessneck Jv, von den Driesch A, Eissa A. Eine Eselsbestattung der 1. Dynastie in Abusir. *Mitteilungen des Deutschen Archäologischen Instituts Abteilung Kairo*. 1992;48:1-10.
6. von den Driesch A. Tierreste aus Buto im Nidelta. *Archaeofauna*. 1997(6).
7. Boessneck J, DRIESCH Avd. Eine Equidenbestattung in spätfrühdynastischer Zeit. *Mitteilungen der Deutschen Orient-Gesellschaft zu Berlin*. 1986(118):45-50.
8. Vila E. *Equids in time and space*. Oxbow Books Oxford; 2006.
9. Zeder M, Meadow R. *Equids in the Ancient World*. Dr. Ludwig Reichert Verlag Wiesbaden, Germany; 1986.
10. Meadow RH. *Equids in the ancient world*. 1986.
11. Uerpmann H-P. *Equids in the Ancient World*. *Naturwissenschaften*. 1991;19:2.
12. Rossel S, Marshall F, Peters J, Pilgram T, Adams MD, O'Connor D. Domestication of the donkey: Timing, processes, and indicators. *Proceedings of the National Academy of Sciences*. 2008;105(10):3715-20.

13. Beja-Pereira A, England PR, Ferrand N, Jordan S, Bakhiet AO, Abdalla MA, et al. African origins of the domestic donkey. *Science*. 2004;304(5678):1781-.
14. Kimura B, Marshall FB, Chen S, Rosenbom S, Moehlman PD, Tuross N, et al. Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society B: Biological Sciences*. 2011;278(1702):50-7.
15. Green MR, Hughes H, Sambrook J, MacCallum P. *Molecular cloning: a laboratory manual*. Molecular cloning: a laboratory manual 2012. p. 1890-.
16. Huang J, Zhao Y, Bai D, Shiraigol W, Li B, Yang L, et al. Donkey genome and insight into the imprinting of fast karyotype evolution. *Scientific reports*. 2015;5:14106.
17. Nilsson MA, Arnason U, Spencer PB, Janke A. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. *Gene*. 2004;340(2):189-96.
18. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-100.
19. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
20. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nature Reviews Genetics*. 2012;13(9):667-72.
21. Pickrell J, Pritchard J. Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*. 2012:1-.
22. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS genet*. 2006;2(12):e190.
23. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*. 2009;19(9):1655-64.
24. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984:1358-70.
25. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-8.
26. Tetens J, Widmann P, Kühn C, Thaller G. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. *Animal genetics*. 2013;44(4):467-71.
27. Takasuga A. PLAG1 and NCAPG-LCORL in livestock. *Animal Science Journal*. 2016;87(2):159-67.
28. Lindholm-Perry AK, Sexten AK, Kuehn LA, Smith TP, King DA, Shackelford SD, et al. Association, effects and validation of polymorphisms within the NCAPG-LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *BMC genetics*. 2011;12(1):103.
29. Metzger J, Schrimpf R, Philipp U, Distl O. Expression levels of LCORL are associated with body size in horses. *PloS one*. 2013;8(2):e56497.
30. Ardestani SS, Aminafshar M, Maryam MBZB, Banabazi MH, Sargolzaei M, Miar Y. A genome-wide signatures of selection study of Welsh ponies and draft horses revealed five genes associated with

- horse type variation. *Gene Reports*. 2020:100833.
31. Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R, Gondro C. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genetics Selection Evolution*. 2015;47(1):66.
 32. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Human molecular genetics*. 2006;15(5):789-95.
 33. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal genetics*. 2015;46(2):110-21.
 34. Ferenčaković M, Hamzić E, Gredler B, Solberg T, Klemetsdal G, Curik I, et al. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics*. 2013;130(4):286-93.
 35. Silió L, Rodríguez M, Fernández A, Barragán C, Benítez R, Óvilo C, et al. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *Journal of Animal Breeding and Genetics*. 2013;130(5):349-60.
 36. Falconer DS. *Introduction to quantitative genetics*: Pearson Education India; 1996.
 37. Zhang Q, Su R, Shan C, Gao C, Wu P. Non-SMC condensin I complex, subunit G (NCAPG) is a novel mitotic gene required for hepatocellular cancer cell proliferation and migration. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics*. 2018;26(2):269-76.
 38. Fernandes I, Bastien Y, Wai T, Nygard K, Lin R, Cormier O, et al. Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and-independent mechanisms. *Molecular cell*. 2003;11(1):139-50.
 39. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, et al. Many sequence variants affecting diversity of adult human height. *Nature genetics*. 2008;40(5):609-15.
 40. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nature genetics*. 2008;40(5):575.
 41. Soranzo N, Rivadeneira F, Chinappan-Horsley U, Malkina I, Richards JB, Hammond N, et al. Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet*. 2009;5(4):e1000445.
 42. Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010;467(7317):832-8.
 43. Weikard R, Altmaier E, Suhre K, Weinberger KM, Hammon HM, Albrecht E, et al. Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on *Bos taurus* growth and lipid deposition. *Physiological genomics*. 2010;42(2):79-88.
 44. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, et al. A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci.

- Human molecular genetics. 2010;19(11):2303-12.
45. N'Diaye A, Chen GK, Palmer CD, Ge B, Tayo B, Mathias RA, et al. Identification, replication, and fine-mapping of Loci associated with adult height in individuals of african ancestry. *PLoS Genet.* 2011;7(10):e1002298.
 46. Carty CL, Johnson NA, Hutter CM, Reiner AP, Peters U, Tang H, et al. Genome-wide association study of body height in African Americans: The women's health initiative SNP health association resource (share). *Human molecular genetics.* 2012;21(3):711-20.
 47. Sovio U, Bennett AJ, Millwood IY, Molitor J, O'Reilly PF, Timpson NJ, et al. Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet.* 2009;5(3):e1000409.
 48. Horikoshi M, Yaghoobkar H, Mook-Kanamori DO, Sovio U, Taal HR, Hennig BJ, et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nature genetics.* 2013;45(1):76-82.
 49. van der Valk RJ, Kreiner-Møller E, Kooijman MN, Guxens M, Stergiakouli E, Sääf A, et al. A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Human molecular genetics.* 2015;24(4):1155-68.
 50. Eberlein A, Takasuga A, Setoguchi K, Pfuhl R, Flisikowski K, Fries R, et al. Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (NCAPG) gene. *Genetics.* 2009;183(3):951-64.
 51. Snelling W, Allan M, Keele J, Kuehn L, Mcdanel T, Smith T, et al. Genome-wide association study of growth in crossbred beef cattle. *Journal of animal science.* 2010;88(3):837-48.
 52. Setoguchi K, Watanabe T, Weikard R, Albrecht E, Kühn C, Kinoshita A, et al. The SNP c. 1326T> G in the non-SMC condensin I complex, subunit G (NCAPG) gene encoding a p. Ile442Met variant is associated with an increase in body frame size at puberty in cattle. *Animal genetics.* 2011;42(6):650-5.
 53. Setoguchi K, Furuta M, Hirano T, Nagao T, Watanabe T, Sugimoto Y, et al. Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate. *BMC genetics.* 2009;10(1):1-12.
 54. Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, et al. A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PloS one.* 2012;7(5):e37282.
 55. Boyko AR, Brooks SA, Behan-Braman A, Castelhana M, Corey E, Oliveira KC, et al. Genomic analysis establishes correlation between growth and laryngeal neuropathy in Thoroughbreds. *BMC genomics.* 2014;15(1):1-9.
 56. Rubin C-J, Megens H-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences.* 2012;109(48):19529-36.

Figures

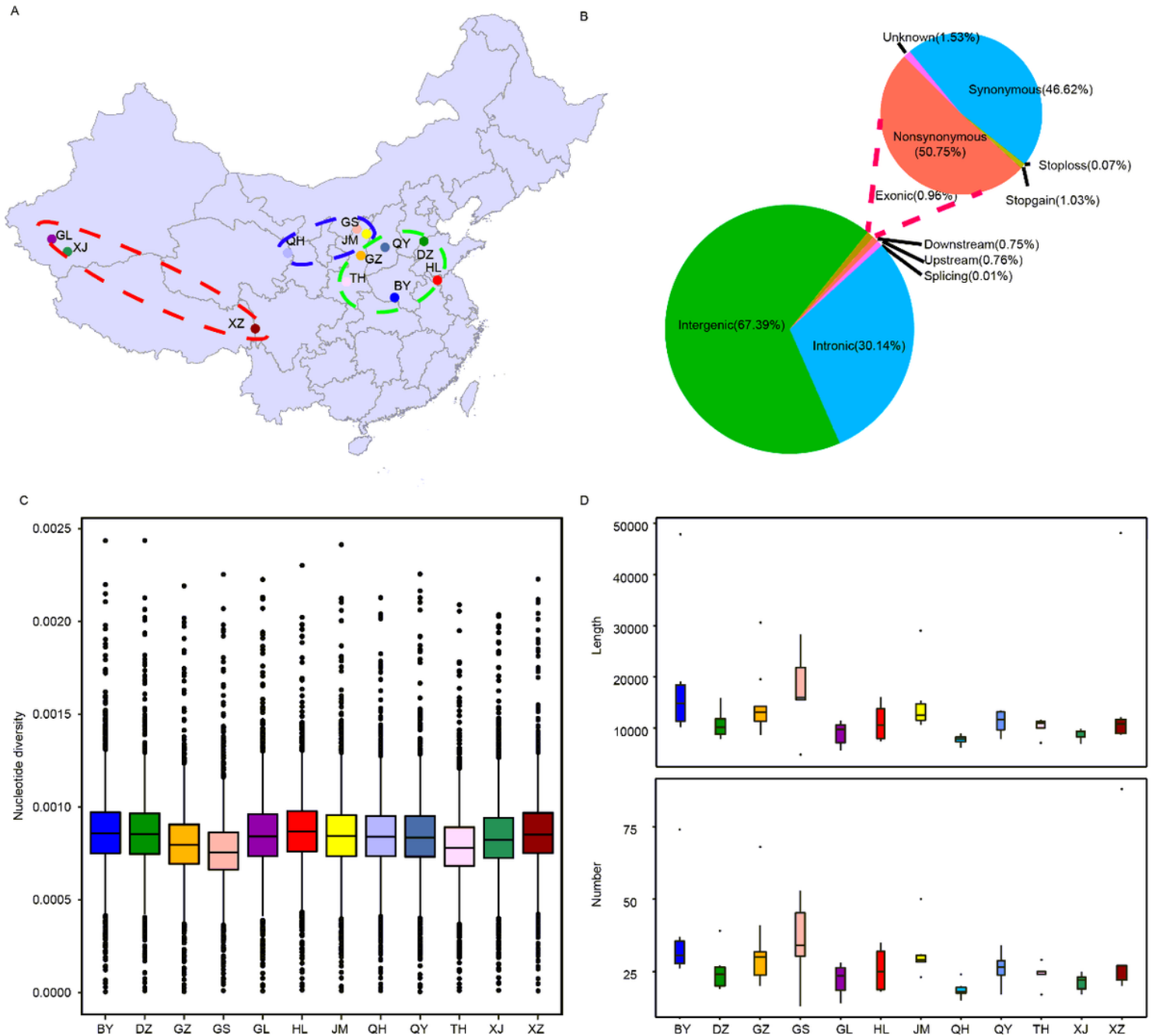


Figure 1

Genomic characteristics of 12 donkey breeds. (A) Sampling sites and geographic distribution of Chinese donkey. Varieties in the green circle represent East China plain donkey, blue circle represents loess plateau donkey, and red circle means Southwest China plateau donkey. (B) Functional classification of detected SNPs. (C) Box plots of nucleotide diversity for each breed in 1-Mb windows with 500-Kb steps. The box indicates the first and third quartiles, and points show outliers. (D) Box plots of ROH length and ROH number for each donkey breed. The box indicates the first and third quartiles, and points show outliers. Note: The designations employed and the presentation of the material on this map do not imply the

expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

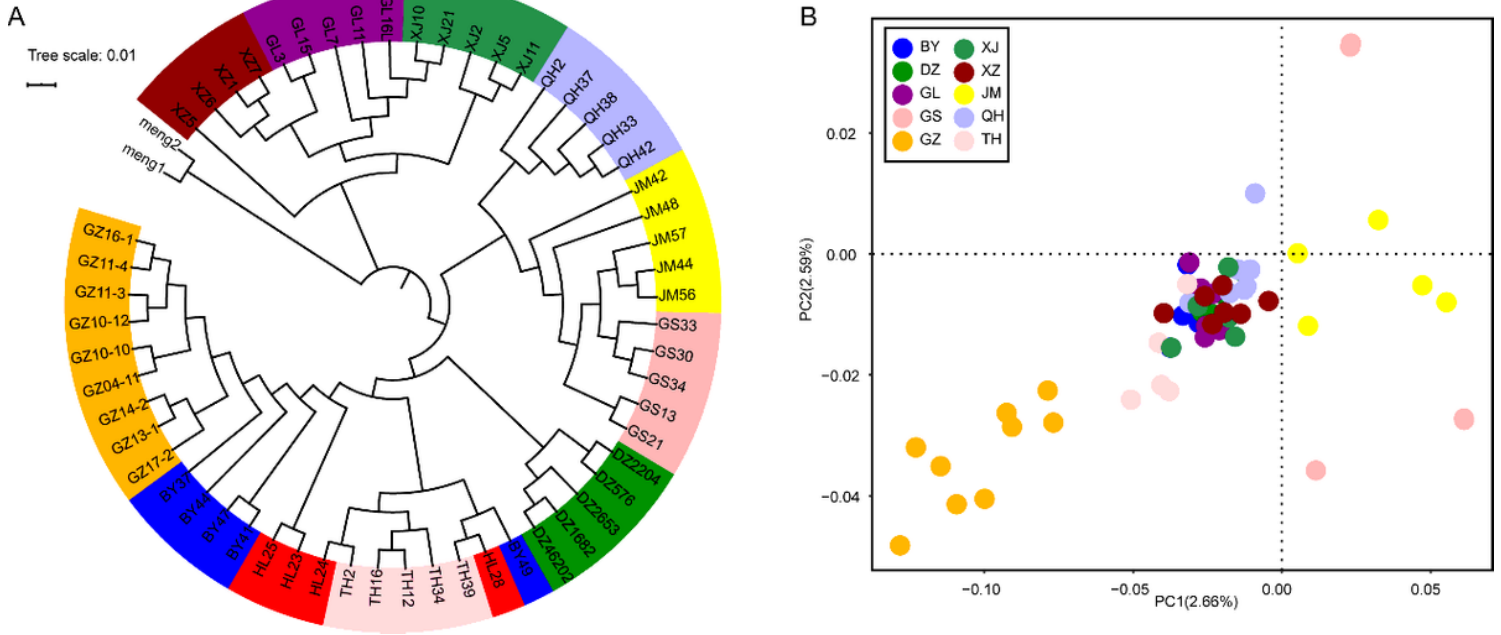


Figure 2

Population genetic structure of the donkey populations. (A) Neighbor-joining phylogenetic tree of 78 Chinese donkeys. Different breeds are shown as different colors. (B) Plots of principal components 1 and 2 from PCA analysis of 78 Chinese donkeys using the whole genome sequencing data.

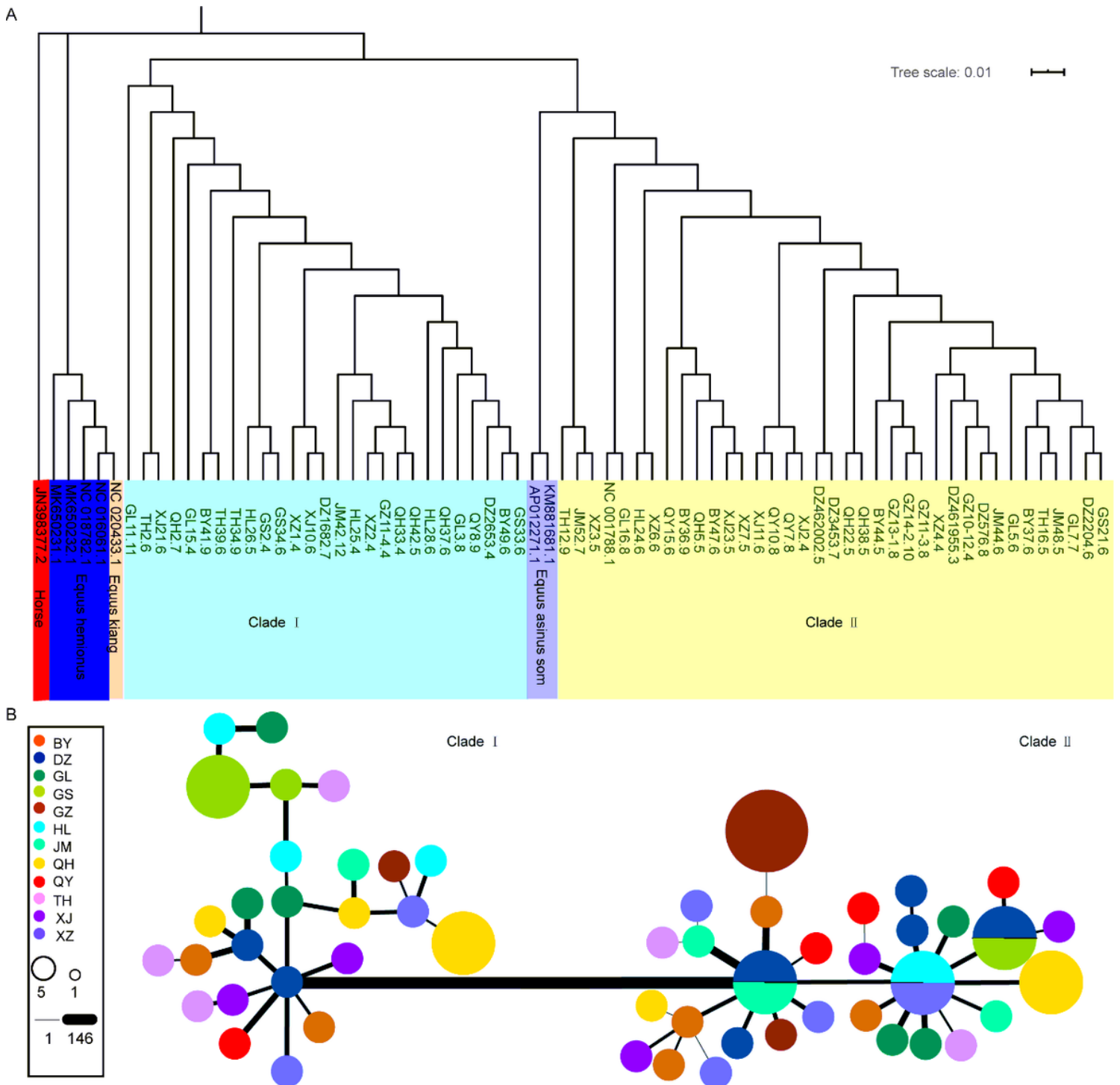


Figure 3

mtDNA genome phylogeny of Donkey. (A) Neighbor-joining tree for Equus. Red: Horse; blue: *E. hemionus*; green: Clade I; purple: *E. asinus somali*; yellow: Clade II. (b) Median-joining network based on mtDNA genome. Different color circles on the left represent different varieties, and circle size is proportional to frequencies of haplotypes.

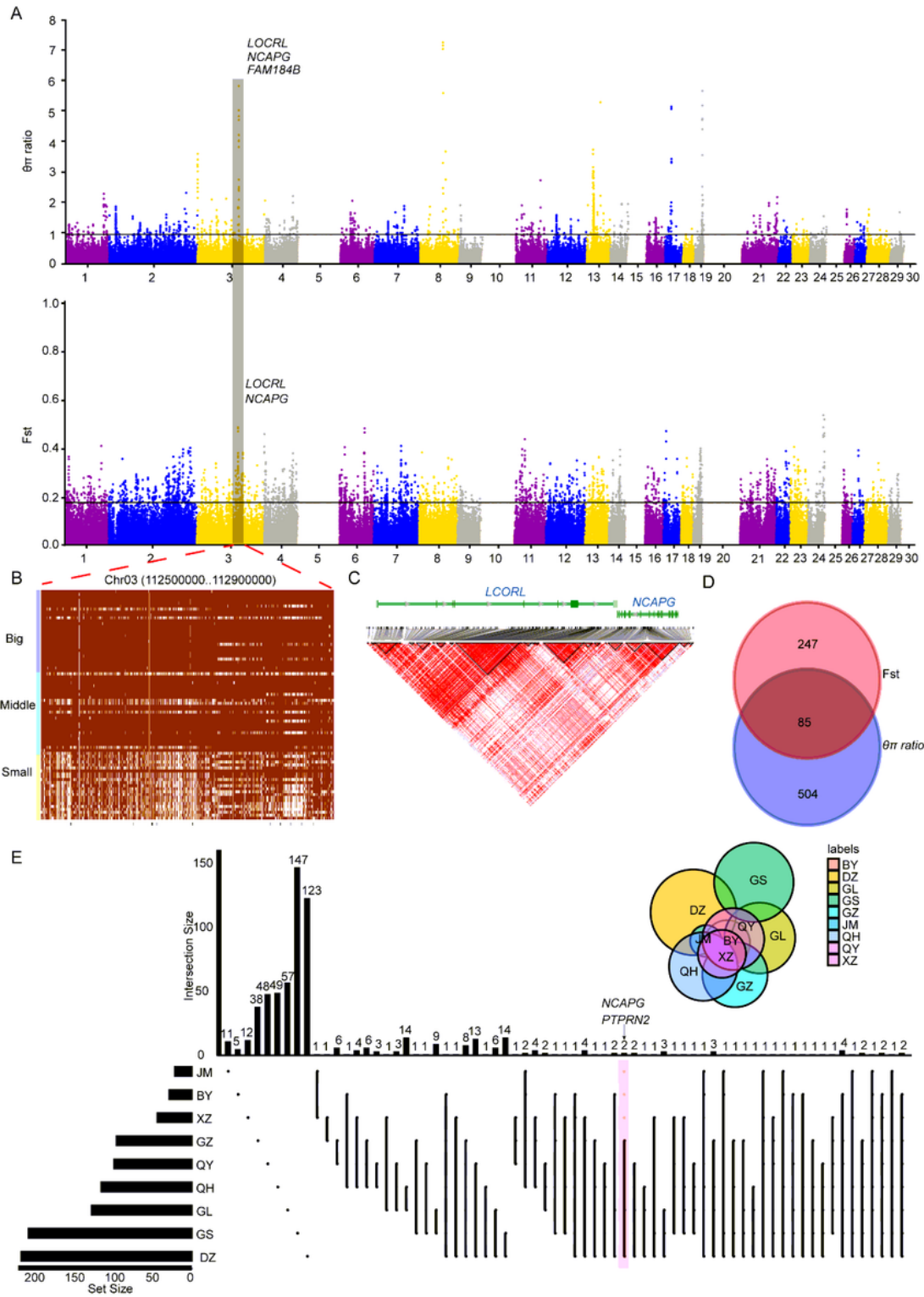


Figure 4

Selective signals of body size among large and small donkey groups. (A) Distribution of the $\theta\pi$ ratio and pairwise fixed index (FST) in a 50-kb sliding window with 20-kb step across all autosome. The black line indicates the threshold ($\theta\pi$ ratio > 0.97 and FST > 0.17). (B) Haplotype diversity in the region of *NCAPG* and *LCORL*. The major allele at each SNP position in large donkeys is colored as deep red, the minor

allele is colored as white. (C) Linkage disequilibrium analysis of the SNPs at the region of NCAPG and LCORL. (d) Venn diagram for 921 genes. (e) Venn diagrams of SweepD results for 9 donkey breeds.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplemental.docx](#)