

Mitogenomic Diversity and Maternal Origins of Yak (*Bos Grunniens*)

Xingdong Wang

Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences

Jie Pei

Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences

Pengjia Bao

Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences

Mengli Cao

Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences

Shaoke Guo

Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences

Rende Song

Animal Disease Prevention and Control Center of Yushu Tibetan Autonomous Prefecture

Weiru Song

Animal Disease Prevention and Control Center of Yushu Tibetan Autonomous Prefecture

Chunnian Liang

Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences

Ping Yan

Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences

Xian Guo (✉ guoxian@caas.cn)

Key Laboratory of Yak Breeding Engineering of Gansu Province, Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences, Lanzhou 730050, China.

Research Article

Keywords: Matrilineal origin, mtDNA, phylogenetic relationship, Yak

Posted Date: February 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-157770/v1>

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

[Read Full License](#)

Mitogenomic Diversity and Maternal Origins of Yak (*Bos grunniens*)

Xingdong Wang^{1,2}, Jie Pei¹, Pengjia Bao¹, Mengli Cao¹, Shaoke Guo¹, Rende Song³, Weiru Song³,
Chunnian Liang¹, Ping Yan^{1*} and Xian Guo^{1*}

¹Key Laboratory of Yak Breeding Engineering of Gansu Province, Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences, Lanzhou 730050, China.

²Life Science and Engineering College, Northwest University for Nationalities, Lanzhou 730030, China.

³Animal Disease Prevention and Control Center of Yushu Tibetan Autonomous Prefecture, Qinghai Province, Yushu 815000, China.

*Corresponding authors: Xian Guo (email: guoxian@caas.cn), Ping Yan (email: pingyanlz@163.com)

ABSTRACT

Background and aim: Mitochondrial genome has the characteristics of simple structure, no recombination, maternal inheritance, high conservation, fast evolution, and high copy number. It is easy to sequence, contains high-resolution phylogenetic information, and exists in a wide range of taxa. Therefore, it is widely used in the study of biological phylogeny. At present, phylogenetic studies focus mainly on D-loop region, cytochrome b gene, and protein coding sequence; phylogenetic studies using the mitochondrial complete sequence are rare. Therefore, this study aimed to conduct phylogenetic experiments using the mitochondrial complete sequence and compare the results with previous findings obtained using partial sequences.

Results: Complete mitochondrial sequences of five yak populations from Qinghai and Xinjiang were obtained. The haplotype diversity of the five populations were Xueduo yak (0.992 ± 0.015), Pamir yak (0.990 ± 0.014), Yushu yak (0.963 ± 0.033), Qilian yak (0.948 ± 0.036), and Huanhu yak (0.905 ± 0.048), which showed a higher haplotype diversity compared with other breeds. A total of 78 haplotypes were obtained from 111 individuals. Among these, Yushu yak, Huanhu yak, Xueduo yak, and Qilian yak all shared haplotypes, but the Pamir yak did not share haplotypes with these four populations. Phylogenetic analysis showed that yak populations were separable into three distinct branches. The analysis identified a new phylogenetic branch with both wild and domestic yaks being represented in the new branch. The 155 haplotypes were divided into 6 haplotype groups by haplotype clustering. The haplotype group had nothing to do with the morphological group of yaks, and yaks from the same population or the same ecological group were distributed in different haplotype groups. The four haplotype groups A, B, C, and D, showed a star-shaped distribution of haplotypes. The central haplotypes were widely distributed and had a high frequency.

Conclusions: The genetic diversity of yaks in Qinghai was high. Both domestic and wild yaks exhibited three different maternal origins. Three different maternal origins existed in ancient wild yaks before domestication, with haplogroup A being the primary haplogroup associated with yak domestication.

Keywords: Matrilineal origin; mtDNA; phylogenetic relationship; Yak

Introduction

Yak (*Bos grunniens*) is a large animal with a compact body, an absence of functioning sweat glands, and a relatively small skin surface area per unit of body weight [1]. These characteristics make yaks well adapted to low-temperature hypoxic environments where they are endemic, such as the QTP Qinghai–Tibetan Plateau (QTP) and adjacent high-altitude regions [2]. China has more than 16 million yaks [3], accounting for more than 95% of the global yak population [4]. Yaks are multipurpose, high-altitude bovid species [5], providing an essential source of economically valuable products for local herders such as milk, meat, fur, and fuel [6]. In addition, yaks provide locals with a means of transportation [7]. Yak thus play an important socioeconomic role in regions where they are endemic, being vital in maintaining pasture ecosystems and agricultural biodiversity in plateau areas [8]; hence, yaks are referred to as "all-around animals" [2]. However, yak populations are currently undergoing degradation, leading to an urgent need to cultivate new breeds. Yaks may have been domesticated from wild yaks by ancient Qiang people in northwest China during the early Holocene period [7] and are the only large animal to live in an environment similar to that of their wild ancestors [9]. Qinghai province is located in the northern QTP and is home to abundant yak genetic resources. Archeological analyses and assessments of mitochondrial DNA (mtDNA) sequences suggest that Qinghai province may be the origin or site of yak domestication [10]. However, further research is warranted to better clarify the maternal origins, phylogenetic structure, and diversity of domestic yak populations by studying the yak mitochondrial genome.

Genetic diversity is a central facet of biological diversity [11], and variations in DNA sequences are the primary drivers of such diversity. MtDNA is a self-replicating, maternally inherited circular DNA molecule that undergoes rapid evolution and no recombination, and contains high frequencies of polymorphic variants [12]. As the entire mitochondrial genome is inherited as a single unit and mutations accumulate therein [13], mtDNA is considered an ideal tool for performing animal studies, evolution, classification, and population genetic diversity [14].

Previous studies analyzed mtDNA to explore bovine genetic diversity [15]. A separate analysis of mtDNA samples from domestic yaks revealed that all domestic yaks originated from a common wild yak ancestor [15], enabling researchers to tentatively identify two domestic yak migration routes [14]: One originates from the eastern part of the QTP through the Himalayas and the Kunlun Mountains to the Pamir region, and the other originates from the eastern part of the QTP through the South Gobi and the Altai Mountains to Mongolia and Russia. Diversity levels of the domestic yak population are the highest in the QTP region [14]. Most mtDNA studies conducted to date have focused explicitly on the control D-loop region of the mitochondrial genome, making it virtually impossible to clearly distinguish certain important branches in livestock [16, 17]. Recently, researchers have discussed the importance of performing complete mtDNA sequencing because such analyses can yield a detailed genetic map when conducted with a sufficiently large sample size [18].

In combination with previous research results, yaks from Qinghai province, which has the highest yak population diversity in the QTP region and is also the presumed origin of yaks, were selected as the research object and Pamir yak from Xinjiang province as the control group in this study. The complete mitochondrial sequences of Yushu, Qilian, Huanhu, Xueduo, and Pamir populations were obtained by sequencing, and the genetic diversity and interspecific genetic distance of the five populations were analyzed. The phylogenetic relationship of yaks was analyzed using the sequencing data and the existing mitochondrial sequence data in GenBank. This study evaluated the maternal origin of yak at the molecular level and laid a foundation for protecting and effectively utilizing bovine resources.

Results

Analysis of base composition and polymorphic sites of mtDNA

This study assessed genetic variations in 111 complete mtDNA sequences (16321–16325) from 4 yak breeds in Qinghai province and 1 yak breed in Xinjiang province as a means of evaluating the genetic diversity, phylogeny, and maternal origins of these breeds. These analyses found that these five yak breeds all exhibited an identical overall base composition (27.3% T, 25.8% C, 33.7% A,

and 13.2% G). Among these, A+T (61%) was higher than G+C (39%), indicating that A + T was substantially more common than G + C in the total yak mitochondrial gene, displaying an AT bias. A total of 114 variable sites were found in Huanhu yaks (1 singleton variable site and 113 parsimony informative sites), 105 in Pamir yaks (1 singleton variable site and 104 parsimony informative sites), 105 in Qilian yaks (all were parsimony informative sites), 122 in Xueduo yaks (all were parsimony informative sites), and 115 in Yushu yaks (all were parsimony informative sites). Through these sequence analyses across all five populations, 150 variable sites were found, including 2 singleton variable sites and 148 parsimony informative sites.

Haplotype analysis of mtDNA

A total of 78 haplotypes were defined for these 111 samples. Among 78 haplotypes, 13 were shared by different populations (accounting for 16.67% of the total haplotypes) and 65 were specific haplotypes of specific populations (accounting for 83.33% of the total haplotypes). The number and type of haplotypes were different among different populations. Of these, haplotype H2 was the most common, being observed 13 times and represented among Huanhu, Qilian, Xueduo, and Yushu yaks. Of the 78 haplotypes observed in these 5 yak populations, only Pamir yaks in Xinjiang did not share haplotypes with other yak populations; the other 4 yak breeds all had shared haplotypes. For more information about haplotypes, refer to Supplementary Table S2.

Genetic diversity analysis of yak mtDNA

The genetic diversity analysis showed that the haplotype diversity of the five yak populations was 0.981 ± 0.007 , and the nucleotide diversity was 0.00272 ± 0.00019 . These data showed that the genetic diversity of five yak populations was high. The nucleotide diversity value of the Pamir yak population (0.00309 ± 0.00018) was higher than that of the other four populations. The haplotype diversity reached a maximum in the Xueduo yak population (0.992 ± 0.015) and a minimum in the Huanhu yak population (0.905 ± 0.048). The D-loop region was found to be the most variable mtDNA sequence in these yaks. Additional details pertaining to genetic diversity estimates, including variable site number, haplotype number, and nucleotide diversity ($P_i \pm SD$), are shown in Table 1.

Analysis of genetic distance between five yak populations

The results of the genetic distance between five yak groups are shown in Table 2. The genetic distance between any two yak groups was small based on the mitochondrial complete genome sequences. The genetic distance between Pamir yak and Qilian yak, Qilian yak and Xueduo yak both are 0.002, and the genetic distance between other yak groups was 0.003. No genetic distance of zero was found between any two populations, indicating differences in the genetic and breeding development of yaks.

Assessment of mtDNA sequence variations and genetic diversity

Combined with the 111 sequences obtained by sequencing, 95 complete mitochondrial genome sequences of yaks that lived in other provinces were downloaded from GenBank (Supplementary Table S3); 206 complete mitochondrial genome sequences contained 155 haplotype sequences (Supplementary Table S4). Base replacement saturation for 155 haplotype sequences was calculated using DAMBE software. The result is shown in Figure 1. The 155 haplotype sequence base substituents in yaks were not saturated, and therefore phylogenetic trees could be constructed.

A neighbor-joining tree (Fig. 2) was next constructed based on these 155 haplotype sequences. This tree separated yaks into 3 branches, of which branch I was the largest (with 103 haplotypes), accounting for 66.45% of the total haplotypes. Branch II contained 52 haplotypes, accounting for 32.26% of the total haplotypes. Branch III was the smallest; only one domestic and one wild yak were included, accounting for 1.29% of the total haplotypes. Phylogenetic analyses revealed that yak mtDNA was separable into six haplogroups (haplogroups A–F). Haplogroup A was the most frequent; it was also the only haplogroup found in five yak groups, including Yushu yak, Qilian yak, Pamir yak, Xueduo yak, and Huanhu yak. This was followed by haplogroup B, which included yaks from all provinces. In contrast, yaks from Yunnan were not included in haplogroup A, whereas yaks from all other provinces were present therein. The aforementioned experimental results showed that branch I comprising A and B was the main branch of yak. In order to verify the reliability of the results, we constructed a Bayesian phylogenetic tree (Fig. 3), and found that

206 yak mitochondrial complete sequences were also divided into 3 large branches and 6 haplogroups. This proves the reliability of data analysis.

A network diagram of 206 yak individuals (Figure 4) was constructed to study yak phylogeny. This diagram mainly comprised three branches, thus verifying the reliability of the branches in Figure 2. Six haplogroups (A–F) were identified, with haplogroups A, B, C, and D exhibiting a star-shaped phylogenetic relationship, and haplogroups E and F exhibiting a dendritic phylogenetic relationship.

Discussion

As a form of matrilineally inherited genetic material, mtDNA has been widely used to study maternal phylogenetic relationships among mammals [19, 20]. Studies of the evolution and taxonomy of yak populations in China to date have primarily focused on the mitochondrial D-loop region [21] and cytochrome b gene (cytb) sequences [16]. The cytb gene is relatively stable in organisms, and its mutation process is slow in mtDNA. Therefore, it is commonly used to reconstruct phylogenetic relationships above the species level. The D-loop region is often used to study the system relationships at the subspecies level because of its rapid variation in mtDNA [15]. Complete mtDNA sequences yield more data than either of these individual isolated sequences, making them more useful than cytb, D-loop regions, or nuclear genes when evaluating mammalian phylogenetic relationships [22–26]. Four yak breeds in Qinghai province and one yak breed in Xinjiang province were sequenced, evaluating the genetic diversity and matrilineal origins of these breeds by comparing them with one another and with yak mtDNA sequences in the GenBank database to develop a phylogenetic model of Chinese yaks.

In this study, mitochondrial full-length sequencing of 111 individuals from 5 yak populations showed that the A + T content was higher than the G + C content, showing the characteristics of mitochondrial base preference. By referring to the literature published in the journal *Mitochondrial DNA Part B*, a general AT preference was found in the yak mitochondrial genome. Specific species often form a set of favorable codon usage patterns in the long-term evolutionary process to adapt to environmental changes [27]. The codon preference is influenced by many

factors, such as genome size, gene sequence length, gene density, gene expression level, CpG island, GC distribution location, tRNA abundance, protein secondary structure, and codon variation preference [28, 29]. Among these, natural selection and mutation are the two main factors affecting codon usage bias, which are also the two factors that have received much attention at present [30, 31]. The base preference of the yak mitochondrial genome may also be to better adapt to the alpine and hypoxic environment at high altitudes. In 111 individuals, 78 haplotypes were found, among which Yushu yak, Huanhu yak, Xueduo yak, and Qilian yak all shared haplotypes, but the Pamir yak did not share haplotypes with these four populations. This might be because the first four populations were close to each other and exchanged genes between populations, while the Pamir yak was far away from the other four species, creating geographical isolation. Further, the haplotype specificity of the Pamir yak was caused by the accumulation of mutations during migration from its origin to the Pamir region.

Genetic diversity is an integral part of all biological diversity. It is the basis of biological evolution and species differentiation, and is of great significance for population maintenance and reproduction and adaptation to habitat changes. The haplotype diversity and nucleotide diversity are important indicators to measure the degree of genetic variation of the population. The higher the haplotype diversity and nucleotide diversity, the higher the degree of genetic variation of the population. The more the genetic diversity, the more likely it is to adapt to different environments. The haplotype diversity of the five populations was higher than that of Jiulong yaks, Maiwa yaks, Zhongdian yaks, Tianzhu white yaks, and Huanhu yaks studied by Zhengchao Tu by sequencing the mitochondrial genomes of 111 yaks from 5 populations in Qinghai and Xinjiang [32]. This was consistent with the result that the diversity level of the yak population was the highest in the QTP [14]. The haplotype diversity (0.905 ± 0.048) of Huanhu yaks in this study was close to that of Huanhu yaks (0.9000) studied by Zhengchao Tu, which verified the reliability of the analysis.

Phylogenetic relationship analyses conducted in this study revealed that yak populations were separable into three distinct branches. Relative to the findings of Lai [16], Guo [33], and Ho [34], the present analysis identified a new phylogenetic branch with both wild and domestic yaks.

However, Wang et al. [35] separated wild yaks into three branches, whereas domestic yaks into two branches. Only a single wild yak was represented in branch III in the present study. In their COIII study of 111 Tibetan yaks, Zhao et al. [36] separated these animals into three branches, consistent with the results of the present study. Previous studies showed highly differentiated intraspecific genetic branches either from several independent domestication events or from a wild species that diverged in an early stage [37, 38]. Together, the findings of this study suggested that yaks might have three matrilineal origins. All three branches were found in both yaks and wild yaks, suggesting that highly differentiated genetic branches had been developed in the early wild yaks. Differences in grouping among studies might be attributable to limited sample sizes in certain analyses, resulting in the overlooking of yaks in the smaller third phylogenetic branch.

The 155 haplotypes were divided into 6 haplotype groups by haplotype clustering, which was consistent with the results of Zhaofeng Wang [35] using mitochondrial protein coding sequence clustering. Zhengchao Tu [32] used the enzyme digestion method to divide yaks into five haplotype groups, revealing that the results of different methods might be inconsistent. The haplotype group had nothing to do with the morphological group of yaks, and yaks from the same population or the same ecological group were distributed in different haplotype groups. Accordingly, each haplotype group contained individuals from different populations or different ecological groups. These results were consistent with the findings of Guo [33] and Wang [35]. The mtDNA of yaks in Qinghai was distributed across haplogroups A–F, but such distribution was not observed for yaks in other provinces. Yaks in Qinghai were also close to the wild yak distribution range, suggesting that Qinghai was likely the site of initial yak domestication [35]. All yaks were divided into six haplotype groups; four of these haplotype groups (A, B, C, and D) showed star-shaped distribution of haplotypes, which was typical of domestic species and consistent with population expansion [39]. The central haplotypes of A, B, C, and D were widely distributed and had a high frequency. Overall, these findings suggested that haplogroup A was associated with the primary yak domestication event. However, the results of this study might be susceptible to haplogroup bias owing to the limited sample size. Additional samples from outside of Qinghai province should be collected to clarify these results in the future.

Conclusions

In summary, the results revealed that the genetic diversity of yaks in Qinghai was high. Both domestic and wild yaks exhibited three different maternal origins, and three different matrilineal origins all existed in ancient wild yaks before domestication, with haplogroup A being the primary haplogroup associated with yak domestication.

Methods

Animals and sample collection

The yaks in Qinghai province, the central producing area of yaks, were selected as the research object, and Pamir yaks in Xinjiang were selected as the control group. A total of 86 yaks in Qinghai province, including 20 yaks in Yushu (N34°7'3", E95°48'18"), 22 yaks in Qilian (N38°11'42", E100°16'45"), 23 yaks in Xueduo (N34°44'30", E101°37'4"), and 21 yaks in Huanhu (N36°55'8", E98°31'19"), were selected. Further, 25 Pamir yaks from Xinjiang province (N38°22'7", E75°47'53") were examined. Venous blood samples were collected from all 111 yaks for mtDNA extraction. The collected blood samples were immediately stored in the in-car refrigerator and transported to the laboratory within 24 h. Then, the blood samples were stored at –80°C in a refrigerator of the Yak Breeding Engineering Laboratory of Gansu province. The storage number of blood samples was R-5-1-001, and DNA was extracted within a month later.

The Animal Care and Use Committee of the Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences approved all yak handling procedures (Permit No: SYXK-2014-0002).

Extraction, amplification, and sequencing of the mitochondrial genome

In the experiment, primers (Table 1, P1–P6) designed by Zhaofeng Wang [35] were used to amplify the entire yak mitochondrial genome, and the primers were sent to Xi'an Qingke Biotechnology Co., Ltd. (Xi'an, China) for synthesis. An EasyPure Blood Genomic DNA Kit (Quanshijin Biotechnology Co., Ltd., Beijing, China) was used to extract DNA from yak blood samples. The extracted DNA was diluted to 50 ng/μL and stored at –80°C until further use. The

DNA concentration and OD260/280 ratio of the samples were determined using a NanoDrop 2000 spectro-photometer (ThermoFisher Scientific, MA, USA). The DNA concentrations and OD260/280 ratios of the samples were 50–1000 ng/μL and 1.6–1.8, respectively. The DNA extraction metrics and tissue-specific metadata are shown in Supplementary Table S1. Ribonucleic acid quality was assessed by evaluating the band on a 1% agarose electrophoretic gel.

Individual polymerase chain reaction (PCR) reactions were conducted using a PCR reaction system, with each reaction containing 2 μL of each primer, 25 μL of 2× ProTaq Master Mix, 2 μL of sample DNA, and 19 μL of ddH₂O. The thermocycler settings were as follows: 95°C for 3 min; 35 cycles of 95°C for 20 s, 52°C for 40 s, and 72°C for 4 min, followed by 72°C for 10 min. The samples were then analyzed by 1% agarose gel electrophoresis to evaluate PCR product size. The samples were then sent to Xi'an Qinco Biotechnology Co., Ltd (Xi'an, China) for sequencing.

Data analysis

MAFFT 7.0 was used for sequence alignment [40], while DNAsp 6.0 [41] was used to calculate the number of variable sites (S), Haplotype diversity (Hd), nucleotide diversity (Pi), average number of nucleotide differences (K), and number of haplotypes (H). DAMBE 7 [42] was used to calculate the saturation of sequence base substitutions. MEGA 6 [43] was used to calculate the frontal genetic distance between five yak populations, and an NJ tree was constructed. Next, 1000 bootstrap replicates were used to assess the reliability of the topology of this tree. A Bayesian phylogenetic tree was constructed using MRBAYES 3.1.2 [44] based on the GTR +G + I Model. Markov Chain Monte Carlo (MCMC) analyses were run for 1 x 10⁷ generations with a burn-in value of 1250. Trees were sampled every 2000 generations. A median-joining network was constructed with NETWORK 10 [45]. The sequences have been deposited in GenBank under the accession numbers MW414100-MW414210.

Table 3. List of abbreviations

Abbreviation	definition
--------------	------------

QTP	Qinghai-Tibetan Plateau
mtDNA	mitochondrial DNA
NJ	neighbor-joining
cytb	cytochrome b gene

Declarations

Ethics approval and consent to participate

All procedures involving animals were performed according to the guidelines of the China Council on Animal Care and the Ministry of Agriculture of the People's Republic of China and approved by the Animal Ethics Committee of Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences.

Consent for publication

Not applicable.

Availability of data and material

The datasets generated and analysed during the current study are available in the GenBank repository, accession number: MW414100-MW414210 (<https://www.ncbi.nlm.nih.gov/genbank/>).

The data supporting the conclusions of this study are available within the supplementary table.

Competing interests

The authors declare no competing interests.

Funding

This study was financially supported by China Agriculture Research System (CARS-37), Innovation Project of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2014-LIHPS-01), and Science and technology plan of Qinghai Province (2020-QY-218; 2019-NK-109).

Authors' Contributions

Conceptualization, X.G. and X.W.; Investigation, R.S. and W.S.; Methodology and software, J.P., S.G., C.L. and X.W.; Data curation, X.W. and P.B.; Writing—original draft, X.W. and M.C.; Writing—review& editing, X.G. and X.W.; funding acquisition, X.G. and P.Y. All authors contributed to the interpretation of the results, and writing of the article.

Acknowledgments

The authors would like to thank all the reviewers who participated in the review and the brothers and sisters for their help in the experiment.

References

1. Krishnan G, Paul V, Biswas TK, Chouhan VS, Das PJ, Sejian V: **Adaptation strategies of yak to seasonally driven environmental temperatures in its natural habitat.** *International Journal of Biometeorology* 2018, **62**(8):1497-1506.
2. Wiener G, Jianlin H, Ruijun L: **The yak.** *Rap Publication* 2011, **44**(4):57-58.
3. Wang H, Chai Z, Hu D, Ji Q, Zhong J: **A global analysis of CNVs in diverse yak populations using whole-genome resequencing.** *BMC Genomics* 2019, **20**(1):61.
4. Jie Z, Zhenzhen X, Ailiang C, Xinyong Y, Yan Z, Wenjing H, Luyao Z, Shuming Y: **Identification of meat from yak and cattle using SNP markers with integrated allele-specific polymerase chain reaction–capillary electrophoresis method.** *Meat Science* 2019, **148**:120-126.

5. Das PP, Krishnan G, Doley J, Biswas TK, Paul V, Chakravarty P, Deb SM, Das PJ: **Identification and expression profiling of MSY genes of yak for bull fertility.** *Journal of Genetics* 2019, **98**(2):1-10.
6. Das PP, Krishnan G, Doley J, Bhattacharya D, Deb SM, Chakravarty P, Das PJ: **Establishing gene Amelogenin as sex-specific marker in yak by genomic approach.** *Journal of genetics* 2019, **98**(1):7.
7. Wang K, Hu Q, Ma H, Wang L, Yang Y, Luo W, Qiu Q: **Genome-wide variation within and between wild and domestic yak.** *Molecular Ecology Resources* 2014, **14**(4):794-801.
8. Shah MA, Xu C, Wu S, Zhao W, Luo H, Yi C, Liu W, Cai X: **Isolation and characterization of spermatogenic cells from cattle, yak and cattleyak.** *Animal Reproduction Science* 2018, **193**:182-190.
9. Zhang X, Wang K, Wang L, Yang Y, Ni Z, Xie X, Shao X, Han J, Wan D, Qiu Q: **Genome-wide patterns of copy number variation in the Chinese yak genome.** *Bmc Genomics* 2016, **17**(1):379.
10. Ma ZJ, Xia XT, Chen SM, Zhao XC, Zeng LL, Xie YL, Chao SY, Xu JT, Sun YG, Li RZ: **Identification and diversity of Y-hromosome haplotypes in Qinghai yak populations.** *Animal Genetics* 2018, **49**(6):618-622.
11. Guo SC, Qi DL, Chen GH, Xu SX, Zhao XQ: **Genetic diversity and classification of domestic yaks inferred from mitochondrial DNA sequences.** *Acta Ecologica Sinica* 2008, **28**(9):4286-4294.
12. Song QQ, Chai ZX, Xin JW, Zhao SJ, Zhong JC: **Genetic diversity and classification of Tibetan yak populations based on the mtDNA COIII gene.** *Genetics & Molecular Research Gmr* 2015, **14**(1):1763-1770.

13. Zeyland J, Wolko u, Lipiński D, Woźniak A, Nowak A, Szalata M, Bocianowski J, Stomski R: **Tracking of wisent–bison–yak mitochondrial evolution.** *Journal of Applied Genetics* 2012, **53**(3):317-322.
14. Ma ZJ, Zhong JC, Han JL, Xu JT, Liu ZN, Bai WL: **Research progress on molecular genetic diversity of the yak (*Bos grunniens*).** *Hereditas* 2013, **35**(2):151-160.
15. Chen PP, Chai ZX, Zhong JC: **Research progress and application of yak mitochondrial gene.** *Heilongjiang Animal Science and Veterinary Medicine* 2015, (19):54-56.
16. Lai SJ, Chen SY, Liu YP, Yao YG: **Mitochondrial DNA sequence diversity and origin of Chinese domestic yak.** *Animal Genetics* 2007, **38**(1):77-80.
17. Tu ZC, Qiu H, Zhang YP: **Polymorphism in mitochondrial DNA (mtDNA) of yak (*Bos grunniens*).** *Biochemical genetics* 2002, **5-6**(40):187-193.
18. Derenko M, Denisova G, Malyarchuk B, Dambueva I, Bazarov B: **Mitogenomic diversity and differentiation of the Buryats.** *Journal of Human Genetics* 2018, **63**(1):71-81.
19. Cai X, Chen H, Lei C: **Matrilineal genetic inter-introgression of *Bos taurus* and *Bos indicus* in China.** *Livestock Science* 2010, **128**(1-3):12-19.
20. Chen S, Lin BZ, Baig M, Mitra B, Lopes RJ, Santos ANM, Magee DA, Azevedo M, Tarroso P, Sasazaki S: **Zebu Cattle Are an Exclusive Legacy of the South Asia Neolithic.** *Molecular Biology & Evolution* 2010, **27**(27):1.
21. Ma Z, Zhong J, Han J, Xu J, Zhang J: **Genetic diversity and demographic history of wild Yak (*Bos grunniens mutus*) inferred from mtDNA D-loop sequences.** *African Journal of Biotechnology* 2010, **9**(46):7805-7810.

22. Mipam TD, Wen YL, Fu CX, Li SR, Zhao HW, Ai Y, Li L, Zhang L, Zou DQ: **Maternal Phylogeny of a Newly-Found Yak Population in China.** *International Journal of Molecular Sciences* 2012, **13**(9):11455-11470.
23. Anderson, T A, Bankier, G B, Barrell, H M, Bruijn D, R A, Coulson, Drouin: **Sequence and organization of the human mitochondrial genome.** *Nature* 1981, **290**(5806):457-465.
24. Gu Z, Zhao X, Li N, Wu C: **Complete sequence of the yak (*Bos grunniens*) mitochondrial genome and its evolutionary relationship with other ruminants.** *Molecular Phylogenetics & Evolution* 2007, **42**(1):248-255.
25. Nishibori M, Hanazono M, Yamamoto Y, Tsudzuki M, Yasue H: **Complete nucleotide sequence of mitochondrial DNA in White Leghorn and White Plymouth Rock chickens.** *Animal Science Journal* 2003, **74**(5):437-439.
26. Xu X, Arnason U: **The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region.** *Gene* 1994, **148**(2):357-362.
27. Li P, Bai YF, Feng RY, Wang YY, Zhang WF: **Analysis of codon bias of NAD-ME gene in *Amaranthushypochondriacus*.** *J Appl Environ Biol* 2011, **17**(01):12-17.
28. Hanson G, Collier J: **Codon optimality, bias and usage in translation and mRNA decay.** *Nature Reviews Molecular Cell Biology* 2018, **19**(1):20-30.
29. Paul P, Malakar AK, Chakraborty S: **Codon usage and amino acid usage influence genes expression level.** *Genetica* 2018, **146**(1):53-63.
30. Sharp PM, Emery LR, Zeng K: **Forces that influence the evolution of codon bias.** *Philos Trans R Soc Lond B Biol* 2010, **365**(1544):1203-1212.
31. Yousheng R, Guozuo W, Zhangfeng W, Xuewen C, Qinghua N, Xiquan Z: **Mutation Bias is the Driving Force of Codon Usage in the *Gallus gallus* genome.** *DNA Research* 2011, **18**(6):499-512.

32. Tu ZC, Zhang YP, Qiu H: **Mitochondrial DNA Polymorphism and Genetic Diversity in Chinese Yaks**. *Chinese Journal of Genetics* 1998, **25**(03):205-212.
33. Guo S, Savolainen P, Su J, Zhang Q, Qi D, Zhou J, Zhong Y, Zhao X, Liu J: **Origin of mitochondrial DNA diversity of domestic yaks**. *BMC Evolutionary Biology* 2006, **6**(1):73.
34. Ho SYW, Larson G, Edwards CJ, Heupink TH, Lakin KE, Holland PWH, Shapiro B: **Correlating Bayesian date estimates with climatic events and domestication using a bovine case study**. *Biology Letters* 2008, **4**(4):370-374.
35. Wang Z, Shen X, Liu B, Su J, Yonezawa T, Yu Y, Guo S, Ho SYW, Vilà C, Hasegawa M *et al*: **Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal**. *Journal of Biogeography* 2010, **37**(12):2332-2344.
36. Zhao SJ, Chen ZH, Ji QM, Chai ZX, Zhang CF, Xin JW, Zhong JC: **Sequence Analysis of mtDNA COIII of Tibetan Yaks**. *Scientia Agricultura Sinica* 2011, **44**(23):4902-4910.
37. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E *et al*: **Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication**. *Science* 2005, **307**(5715):1618-1621.
38. Pedrosa S, Uzun M, Arranz JJ, Gutierrez-Gil B, San Primitivo F, Bayon Y: **Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events**. *Proceedings of the Royal Society B: Biological Sciences* 2005, **272**(1577):2211-2217.
39. Sun T, Huang G, Sun J, Wang Z, Teng S, Cao Y, Hanif Q, Chen N, Lei C, Liao Y: **Mitogenome Diversity and Maternal Origins of Guangxi Buffalo Breeds**. *Animals* 2020, **10**(4):547.

40. Katoh K, Rozewicki J, Yamada KD: **MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization.** *Briefings in Bioinformatics* 2019, **20**(4):1160-1166.
41. Julio R, Albert FM, Carlos S-DJ, Sara GR, Pablo L, E R-OS, Alejandro S-G: **DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets.** *Molecular Biology & Evolution* 2017, **34**(12):3299-3302.
42. Xuhua X: **DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution.** *Molecular Biology and Evolution* 2018, **35**(6):1550-1552.
43. Koichiro, Tamura, Glen, Stecher, Daniel, Peterson, Alan, Filipowski, Sudhir, Kumar: **MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0.** *Molecular biology and evolution* 2013, **30**(12):2725-2729.
44. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**(12):1572-1574.
45. Bandelt HJ, Forster P, Röhl A: **Median-joining networks for inferring intraspecific phylogenies.** *Molecular Biology & Evolution* 1999, **16**(1):37-48.

Figure legends

Figure 1. Test of base substitution saturation for 155 haplotype sequences. Note: S: conversion; V: transversion; $S/V > 1$ means that the sequence base replacement is not saturated.

Figure 2. A NJ tree of 155 yak haplotypes. The reliability of the topology of this tree were used 1000 bootstrap replicates.

Figure 3. A mitogenome network of 206 yaks. The red dot in the figure represents wild yak, and the yellow dot represents domestic yak.

Table 1. Genetic structure and diversity of Qinghai and Pamir yaks

Breed	N	Haplogroup						S	H	K	Hd±SD	Pi±SD
		A	B	C	D	E	F					
Huanhu	21	3	3	1	0	0	0	114	12	35.943	0.905±0.048	0.00220±0.00061
Pamir	25	10	1	2	7	1	1	105	22	50.447	0.990±0.014	0.00309±0.00018
Qilian	22	8	0	2	3	0	1	105	16	44.610	0.948±0.036	0.00273±0.00043
Xueduo	23	12	2	2	1	0	0	122	21	37.921	0.992±0.015	0.00232±0.00051
Yushu	20	8	2	3	1	1	1	115	16	45.005	0.963±0.033	0.00276±0.00047
Total	111	41	8	10	12	2	3	150	78	44.470	0.981±0.007	0.00272±0.00019

N: number of yaks; S: number of variable sites; H: number of haplotypes; K: the average number of differences; Hd: haplotype diversity; Pi: nucleotide diversity; SD: standard deviation

Table 2. Complete mitochondrial sequence genetic distance between five yak populations

Breed	Huanhu	Pamir	Qilian	Xueduo
Pamir	0.003			
Qilian	0.002	0.003		
Xueduo	0.002	0.003	0.003	
Yushu	0.003	0.003	0.003	0.003

Figures

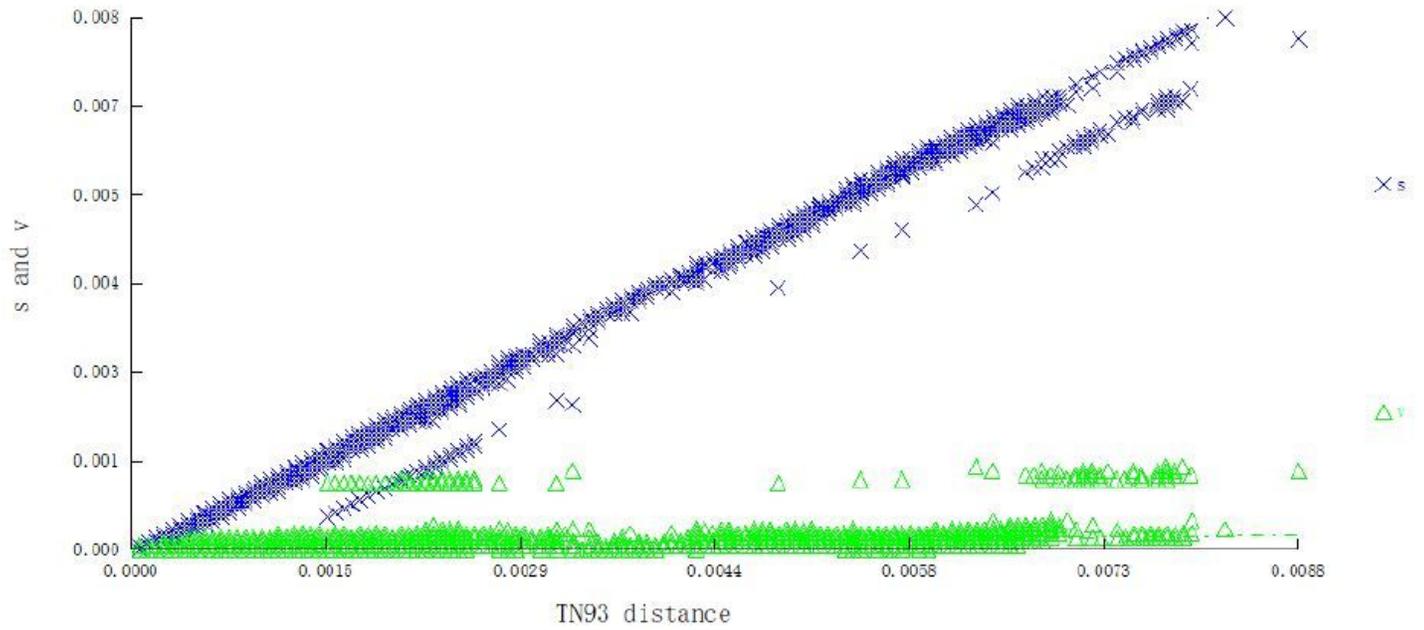


Figure 1

Test of base substitution saturation for 155 haplotype sequences. Note: S: conversion; V: transversion; $S/V > 1$ means that the sequence base replacement is not saturated.

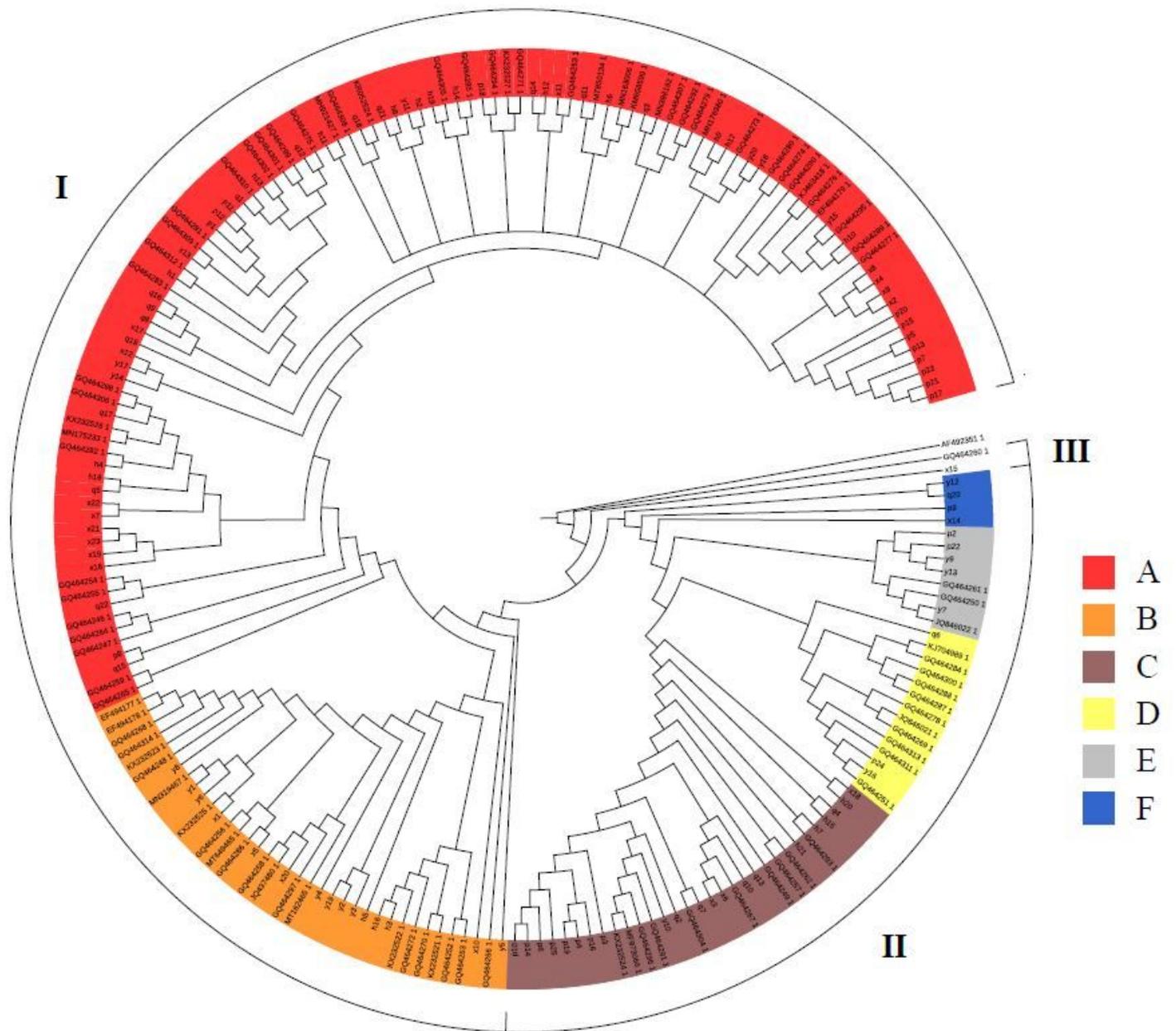


Figure 3

A mitogenome network of 206 yaks. The red dot in the figure represents wild yak, and the yellow dot represents domestic yak.

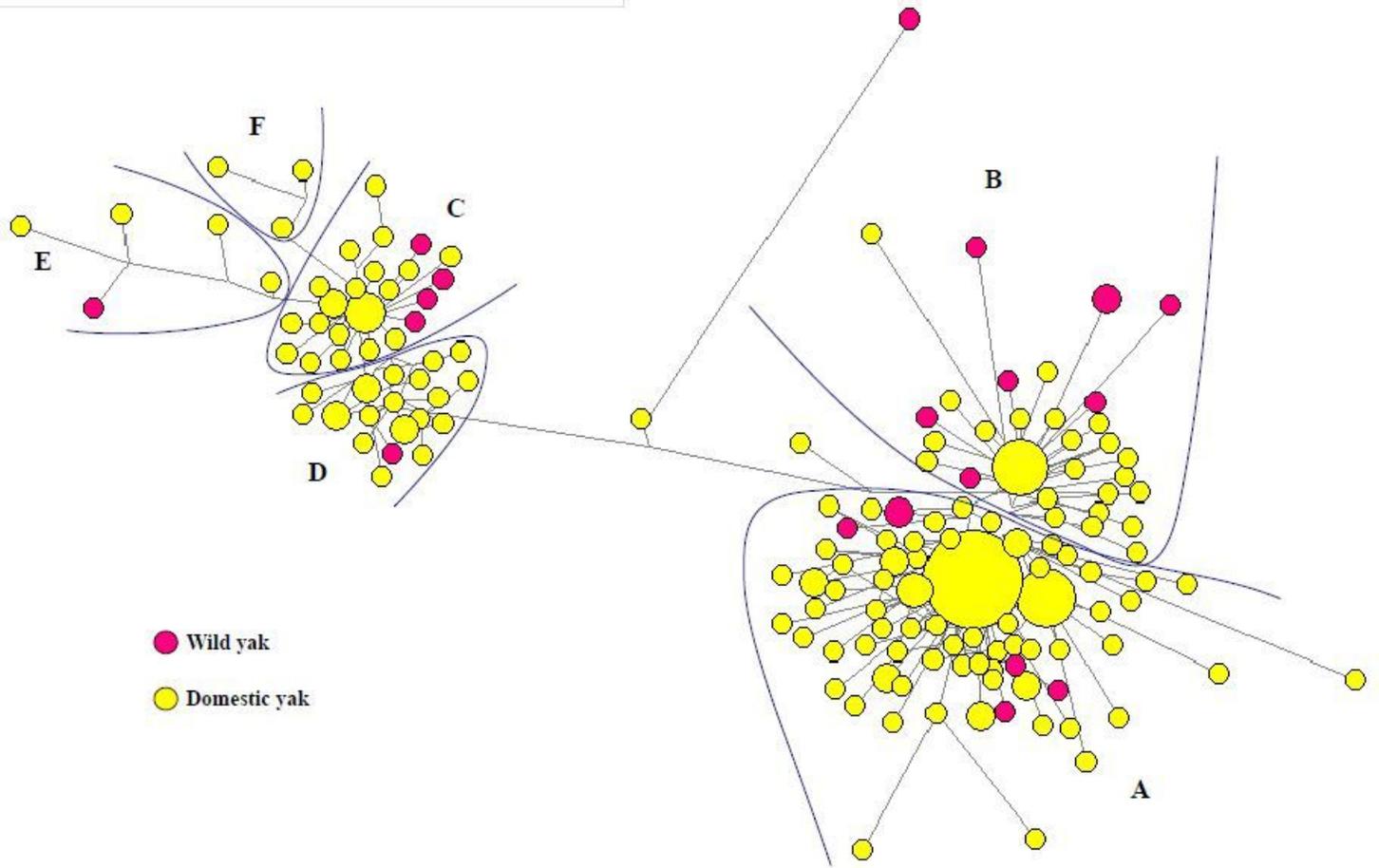


Figure 4

A network diagram of 206 yak individuals (Figure 4)

Supplementary Files

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

- [supplementarytable.docx](#)