Genetics and Molecular Research

Genetic diversity and population structure of Kazakh horses (*Equus caballus*) inferred from mtDNA sequences

M. Gemingguli[1], K.R. Iskhan[2], Y. Li[3], A. Qi[4], W. Wunirifu[5], L.Y. Ding[1] and A. Wumaierjiang[1]

[1]College of Animal Science, Tarim University, Alar, Xinjiang, China
[2]Department of Animal Science, Kazakhstan National Agrarian University, Almaty, The Republic of Kazakhstan
[3]Institute of Arid Ecology and Environment, Xinjiang University, Urumqi, China
[4]Genomics Laboratory of Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China
[5]College of Animal Science and Veterinary, Inner Mongolia Agricultural University, Huhhot, China

Corresponding author: A. Wumaierjiang
E-mail: wumaier@tom.com

Genet. Mol. Res. 15 (4): gmr.15048618
Received March 10, 2016
Accepted July 5, 2016
Published October 5, 2016
DOI http://dx.doi.org/10.4238/gmr.15048618

ABSTRACT. The Kazakh horse is an important old horse breed in Xinjiang. They have contributed greatly to the breeding and improvement of other local horse breeds, yet their genetic diversity and population structure are not well understood. In the present study, we evaluated the genetic diversity of Kazakh horses and their relationship with other horse breeds using the mtDNA D-loop region, *Cyt b* gene, and a DNA fragment (nps 7974-9963, containing *COX3*, *tRNA-Gly*, *ND3*, and *tRNA-Arg*). A total of 130 Kazakh horses from

8 populations in China and Kazakhstan were analyzed. A total of 88 haplotypes (haplotype diversity: 0.9895) were identified, in which 3 haplotypes were shared by groups in the two countries. In a median-joining network, 6 haplogroups were found, in which most haplogroups included haplotypes from different populations. Neighbor-joining analysis revealed similar results in that haplotypes in different populations were admixed in most of the 6 clusters. In conclusion, a high level of genetic diversity was found in the Kazakh horses. However, no clear correspondence between haplogroups and geographic origin and no significant differentiation between populations in the two countries were observed. This might have resulted from the frequent contact between the two countries through the Silk Road in the past, or due to long-term outcrossing and hybridization with the introduced horses.

Key words: Kazakh horses; Genetic structure; Population diversity; D-loop; *Cyt b*

INTRODUCTION

An understanding of the genetic characteristics of a population is essential to implement conservation strategies that preserve genetic diversity (Hall and Bradley, 1995). Appropriate assessment and characterization of populations at the phenotypic and genotypic levels are some of the first and crucial steps in the development of conservation strategies (Plante et al., 2007; Gizaw et al., 2008). Molecular data can aid in identifying animals or sets of animals that should be preserved to prevent loss of genetic diversity. Based on genetic diversity and population structure analysis of Brazilian Mangalarga Marchador horses, DeAssis et al. (2009) demonstrated the low level of genetic diversity of Mangalarga Marchador horses and recommended two conservation strategies: avoidance of crosses between highly endogamic individuals and stimulation of crosses between individuals from those regions identified as having low gene flow. In addition, genetic diversity studies of the Kiso horse indicated that the genetic diversity of the maternal lineage has been reasonably well maintained. However, since the population was small, it might be in the extinction vortex and the genetic diversity of the maternal lineage of Kiso horse would be decreasing. Therefore, key measures, such as the use of reproductive technologies and investigation of new ways to utilize the horses, should be implemented for preservation (Takasu et al., 2014).

The Kazakh horse is an old horse breed. The present Kazakh horses are thought to be the descendants of the ancestral Wusun horses of the 2nd century B.C. The development of the present Kazakh horse is the result of long-term breeding and improvement. Their development are closely related to the need for nomadic transportation along the Silk Road, wars, and horse exchanges between China and Central Asia. Kazakh horses are mainly distributed in western China, Kazakhstan, northeast Kyrgyzstan, Mongolia, and the western Altay region of Russia. In China, they are mainly found in the areas of the Tianshan Mountains, west of Junggar Basin, and the western part of the Altay Mountains. The greatest numbers of Kazakh horses are in Xinyuan County, while Nilka County has the greatest numbers of purebred horses. Kazakh horses in Zhaosu, Turks, and Gongliu County were selected and bred to form the Ili horse. In addition to the Ili region, Boertala Mongol Autonomous Prefecture, Tacheng, the

Altay region, and Changji Hui Autonomous Prefecture also raise Kazakh horses, but they are generally fed together with Mongolian horses (Zhao, 1991).

The typical characteristics of Kazakh horses include moderately large head, short ears, slender and slightly raised neck, high withers, narrow chest, and often knife-like hind legs. The length and height of adult horses average 150 and 152 cm and 145 and 135 cm for stallions and mares, respectively. They are well adapted to the cold climate, moderate in temperament, and grow rapidly (Gemingguli, 2011). An adult horse (4-12 years) weighs 340 to 440 kg and the meat production rate is around 57%. A mare can produce 5 to 6 kg milk per day under grazing conditions. The meat and milk are delicious and rich in nutrients necessary for humans. They play a key role in the health and lives of local people, as well as in the development of the local economy and society (Tiemuerbai, 2014).

Over a long period of time, because close attention was not paid to breeding, the quality and quantity of Kazakh horses in China dramatically decreased, which led to their smaller physique, weight loss, and other desirable features of the breed gradually degenerating (Tiemuerbai, 2014). The number of Kazakh horses has reduced significantly in recent years with the increase in the consumption of horse meat in the Xinjiang region. Moreover, the number of purebred Kazakh horses is decreasing because the majority of Kazakh horses (such as in the Ili regions) have been hybridized with introduced horses. Therefore, effective measures to conserve and improve this rare horse should be developed and implemented (Gemingguli, 2011).

Owing to the high level of base pair substitution rate compared with nuclear DNA (Brown et al., 1979), strict maternal inheritance (Hutchison et al., 1974), and lack of recombination (Bowling et al., 2000), mtDNA is powerful for clarifying scientific issues regarding horses, such as genetic diversity in Serbia horses (Ocokoljic et al., 2013), Italian horses (Bigi et al., 2014), and Sicilian autochthonous horse breeds (Guastella et al., 2011), adaptive evolution (Ning et al., 2010), phylogenetic relationships between horse populations (Mirol et al., 2002; Prystupa et al., 2012; Winton et al., 2013), and the origin of some horse breeds (Cieslak et al., 2010; Guastella et al., 2011; Lippold et al., 2011; Achilli et al., 2012).

Several studies have been conducted on horses using mtDNA; however, only few studies have focused on Kazakh horses. To date, little is known about their diversity or phylogenetic relationships with other horse breeds. The aims of this study were to assess the genetic diversity and population structure of the Kazakh horses using mtDNA sequences, including the D-loop region, the cytochrome oxidase *b* (*Cyt b*) gene, and a DNA fragment (nps 7974-9963, containing gene partial sequences for the *COX3* gene, *tRNA-Gly*, the *ND3* gene, and *tRNA-Arg* sequences). Phylogenetic relationships between Kazakh horses and other breeds worldwide were also analyzed.

MATERIAL AND METHODS

Venous blood (2-3 mL) was drawn from the neck of each adult horse and immediately stored in EDTA-containing anticoagulation vacuum blood vessels at -20°C. A total of 130 samples were collected from 130 Kazakh horse individuals from 8 populations; 70 individuals were from 6 populations: Altai (AL), Hefeng (HE), Tuoli (TE), Bole (BL), Hemu (HM), and Manasi (MN) in China and 60 individuals were from 2 populations: Taldykorgan (TAL) and Almaty (ALM) in Kazakhstan. The treatment of horses followed the supporting institution's ethical standards. In addition, 168 accessions of D-loop sequences from 25 horse breeds were retrieved from GenBank to investigate the relationships between Kazakh horses and other horse breeds (**Table S1**).

Total genomic DNA was extracted using a rapid blood genomic DNA extraction kit (Sangon Biotech, China). The primers used for amplification of the *Cyt b* gene, D-loop region, and a DNA fragment (nps 7974-9963 containing the *COX3* gene, *tRNA-Gly*, *ND3* gene, and *tRNA-Arg* sequences, hereafter called the 7974 fragment) are listed in Table 1. PCR primers were synthesized by Sangon Biotech. PCRs included Qiagen HotStar HiFidelity polymerase (Germany). Each reaction was performed in a volume of 25 μL using 13.5 μL ddH_2O, 5 μL HotStar PCR buffer, 2.5 μL forward primer (10 μM), 2.5 μL reverse primer (10 μM), 1 μL (30-50 ng) DNA sample, and 0.5 μL HotStar polymerase. DNA amplifications were performed in a Gene-Amp PCR system 9700 (Applied Biosystems, USA) under the following conditions: initial denaturation at 95°C for 5 min; followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 1.5 min; and a final extension at 72°C for 10 min. PCR products were purified with a TIAN Quick Midi Purification Kit (TIANGEN, Beijing, China) and sequenced bidirectionally with forward and reverse PCR primers using an ABI PRISM™ 3730 DNA sequencer by Sangon Biotech.

Table 1. Primers used for DNA amplification in this study.

Gene		Primer sequence (5'-3')
Cyt b	F	TCATTATTCCCACGTGGAATCTAAC
	R	GATAGTCCTTGGGAGGAAACGTT
D-loop	F	CAAGGACTATCAAGGAAGAAGCTCT
	R	TGGAGTAAGAATACTCATCTAGGCA
7974 fragment	F1	TCTCAAAGCACTCCTATCCGACAC
	F2	CAGCCTATTTATTCAACCTGTAGC
	F3	TTCTGAGCCTTTTACCACTC
	F4	TGACTTCCAATCAATCAGCT
	R	CTCGTTCGCAGGCAGCAAAT

All sequences were examined by eye in Chromas 2.4 and BioEdit version 7.0.5.3. All sequences were initially aligned using the Clustal W software (Thompson et al., 1994) with default settings. The statistical quantities for the DNA sequences, including number of haplotypes (Nh), haplotype diversity (Hd), and nucleotide diversity (π), were calculated using DnaSP 5.10.1 (Librado and Rozas, 2009). The genetic distances within and between populations were calculated with the MEGA 6 software package (Tamura et al., 2013). To represent the genetic structure and differentiation of tested populations, analysis of molecular variance (AMOVA) and pair-wise F_{ST} were conducted. AMOVA and pairwise F_{ST} were performed using the Kimura 2-parameter model with Arlequin v3.5 (Excoffier and Lischer, 2010). *Cyt b*, D-loop, and the 7974 fragment were evaluated independently and in combination. For the interpretation of pairwise F_{ST} results, we followed the suggestion that refers that values between 0-0.05 indicates little genetic differentiation; values between 0.05 and 0.15, moderate differentiation; values between 0.15 and 0.25, great differentiation; and values above 0.25, very great genetic differentiation (Wright, 1978; Hartl and Clark, 1997). A median-joining network (MJ network) was constructed to detect haplotypic differentiation in the horse populations with the NETWORK 4.6 software (Bandelt et al., 1999), based on concatenated combined sequences. Default settings were applied. Within each haplotype in the MJ network, different color codes represented the proportions of different populations in each haplotype.

Phylogenetic analysis of the haplotypes was performed using the neighbor-joining (NJ) and Bayesian method with the MEGA 6 software package (Tamura et al., 2013) based on the Kimura 2-parameter model and the BEAST v1.7 software package (Drummond et al.,

2012) based on the TN93+G+I model. The 50% majority-rule consensus NJ tree was obtained from 1000 bootstrap replications. For BEAST analyses, two independent runs of 10 million generations each were performed, sampling every 1000th generation. Each run was checked in Tracer v.1.5 for sufficient effective sample size (over 100 as suggested by the authors) in the main relevant statistics. The two runs (log- and tree-files) were then combined with LogCombiner v.1.7.4. The maximum clade credibility tree was generated from the 90% post-burn-in trees file using TreeAnnotator 1.7.4. Finally, Figure Tree 1.4.0 was used to display phylogenetic trees.

RESULTS

Genetic diversity

Genetic diversity is usually estimated by using several parameters, e.g., the average number of nucleotide differences (k, mean number of pairwise differences in a DNA dataset of a population), number of polymorphic sites (Nps, number of variable nucleotides regarding one specific DNA region in the sample population), nucleotide diversity (π, a measure of genetic variation, defined as the average number of nucleotide differences per site between any two DNA sequences chosen randomly from the sample population), number of haplotypes (Nh), and haplotype diversity (Hd, a measure of the uniqueness of a particular haplotype in a given population) for each population or group.

For each of the D-loop, *Cyt b* gene, and the 7974 fragment, 130 sequences were obtained. The alignment length was 708, 1140, and 2020 bp, respectively. The nucleotide diversity (π) was 0.020, 0.0048, and 0.0056 for the three DNA regions, respectively. The number of haplotype and haplotype diversity for the three DNA regions were 74/0.985, 28/0.940, and 47/0.959, respectively (Table 2).

After concatenating the *Cyt b*, D-loop, and the 7974 fragment, a sequence with a total of 3868 bp was obtained. Among 222 polymorphic sites, 162 were potentially phylogenetically informative. The k (gaps/missing data were excluded) and π were 25.55 ± 11.3 and 0.0095. The k and π for China horse groups were 25.67 ± 11.39/0.0094 and 25.43 ± 11.31/0.0093 for Kazakhstan groups, respectively. The corrected pairwise difference between China and Kazakhstan groups was 0.02 ($P > 0.01$). A total of 88 haplotypes (Hd: 0.9895) were identified in which 3 haplotypes, H9, H27 and H43, were shared by the two geographic groups. In China populations, 53 unique haplotypes were found and 32 were observed in Kazakhstan populations. The K-2-P genetic distance within the 8 populations ranged from 0.005 to 0.012 and between populations from 0.008 to 0.011 (Table 3).

Distances between populations are below the diagonal and SE values are above the diagonal. Distances within a population (bold) are on the diagonal.

Network analysis and population differentiation

The MJ network resolved 6 haplogroups, which were referred to as groups A, B, C, D, E, and F (Figure 1). Group A included 29 haplotypes, 11 were from Kazakhstan, 20 from China. Haplotypes H9 and H43 were shared by China and Kazakhstan. The 19 haplotypes of B were from Kazakhstan (10) and China (9). The 24 haplotypes in C were also from China (16) and Kazakhstan (9) and 1 shared between them (H27).

Table 2. Populations tested in the study.

Geographic location		D-loop					*Cyt b*					7974 fragment				
	Population	N	Nh	k	Nps	π	N	Nh	k	Nps	π	N	Nh	k	Nps	π
China	AL	13	10	15.4	56	0.022	13	6	4.1	13	0.0036	13	6	10.2	33	0.005
	TE	14	13	14.4	47	0.02	14	8	4.36	17	0.0038	14	12	13.6	56	0.0066
	HE	8	3	6	19	0.0081	8	4	3.93	9	0.0035	8	4	8.29	25	0.004
	BL	7	5	13.4	32	0.019						7	5	8.2	18	0.0041
	HM	16	13	15.5	57	0.022						16	12	11.3	40	0.0056
	MN	13	11	15.3	44	0.021						13	8	10.8	34	0.0054
	Total	71	49	15.0	84	0.021	35	14	4.6	24	0.004	70	33	11.2	76	0.0055
Kazakhstan	TAL	2	2	22	22	0.025	2	2	7	7	0.0061	3	3	13.3	20	0.0066
	ALM	57	28	14.5	69	0.02	57	19	5.9	35	0.0051	57	25	11.3	69	0.0056
	Total	59	30	14.6	71	0.020	59	20	5.9	36	0.005	60	25	11.3	69	0.0056
	Total	130	74 (0.985)			0.020		28 (0.94)			0.0048		47 (0.96)			0.0056

N: number of individuals in each population. Nh: the number of haplotypes in each population and the haplotype diversity (Hd) are given in parentheses. Nps: the number of polymorphic sites. π: nucleotide diversity. k: average number of nucleotide differences.

Table 3. K-2-P genetic distances within and between populations.

Population	AL	HE	TE	BL	HM	MN	TAL	ALM
AL	**0.009**	0.001	0.001	0.001	0.001	0.001	0.001	0.001
HE	0.011	**0.005**	0.001	0.001	0.001	0.001	0.001	0.001
TE	0.010	0.010	**0.010**	0.001	0.001	0.001	0.001	0.001
BL	0.010	0.008	0.009	**0.008**	0.001	0.001	0.001	0.001
HM	0.011	0.009	0.010	0.009	**0.010**	0.001	0.001	0.001
MN	0.010	0.009	0.010	0.009	0.010	**0.010**	0.001	0.001
TAL	0.011	0.008	0.011	0.009	0.011	0.010	**0.012**	0.001
ALM	0.010	0.009	0.010	0.009	0.010	0.010	0.010	**0.010**

Most of the 7 haplotypes in D were from China (6), only 1 from Kazakhstan. Group E had 7 haplotypes in total with 3 from China and 4 from Kazakhstan. Haplotypes from China and Kazakhstan were admixed. There was no congruence of haplogroup distribution to populations from different geographic areas.

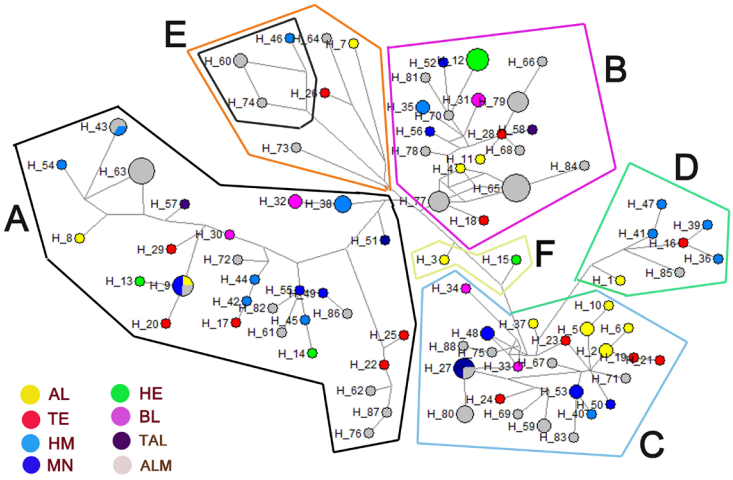

Figure 1. Median-joining network based on combined mitochondrial DNA sequences representing 130 horses within 88 haplotypes. Each population is indicated by a different color and the proportions of different populations within each haplotype are shown. Different haplogroups (**A**-**F**) are indicated.

For AMOVA, we divided the 8 populations into 2 groups according to geographic origins. One group combined populations in China and the other included populations in Kazakhstan. The results partitioned variation in the D-loop to -1.83% (F_{CT} = -0.018) occurring among groups, 6.98% (F_{SC} = 0.069, P < 0.01) among populations within groups, and 94.84% (F_{ST} = 0.052, P < 0.01) occurring within populations. In *Cyt b*, -1.51% of the variation (F_{CT} = -0.015) was explained by differences among groups, 6.36% (F_{SC} = 0.063) among populations within groups, and 95.15% (F_{ST} = 0.049) within sampled populations. For the 7974 fragment sequences, 0.70% (F_{CT} = 0.007) of variation occurred among groups, 1.66% (F_{SC} = 0.017) among groups within populations, and 97.63 (F_{ST} = 0.024) within populations (Table 4). AMOVA results based on combined sequences showed that -0.91% (F_{CT} = -0.009) of variation occurred among groups, 5.35% (F_{SC} = 0.053) among populations within groups, and 95.56% (F_{ST} = 0.044) within populations (Table 4).

Table 4. Analyses of molecular variance based on D-loop, *Cyt b*, and 7974 fragment sequences for the horse populations.

Source of variation	Percentage of variation (D-loop)		Percentage of variation (*Cyt b*)		Percentage of variation (7974)		Percentage of variation (combined)	
Among groups	-1.83	F_{CT} = -0.018	-1.51	F_{CT} = -0.015	0.70	F_{CT} = 0.007	-0.91	F_{CT} = -0.009
Among populations within groups	6.98	F_{SC} = 0.069**	6.36	F_{SC} = 0.063	1.66	F_{SC} = 0.017	5.35**	F_{SC} = 0.053
Within populations	94.84	F_{ST} = 0.052**	95.15	F_{ST} = 0.049	97.63	F_{ST} = 0.024	95.56**	F_{ST} = 0.044

**P < 0.01.

Of 28 pairwise F_{ST} values between the 8 populations based on combined data, 17 comparisons had F_{ST} values between 0 and 0.05 showing little genetic differentiation while 8 comparisons had F_{ST} values between 0.05 and 0.15 showing moderate genetic differentiation but not significant (P > 0.01), and 2 comparisons (AL-HE and HE-TE) had F_{ST} values between 0.15 and 0.25 showing great genetic differentiation. Negative F_{ST} values were recorded in some comparisons and these equate to zero F_{ST} values. While most of the lowest F_{ST} values were seen between populations of TE, BL, HM, and MN and TAL and ALM, the highest F_{ST} values were between the HE population and three other populations (AL, TE, and HM). AL and HE comparisons showed high genetic differentiation, with F_{ST} value of 0.296 (Table 5). The pairwise F_{ST} values between China and Kazakhstan groups were 0.02 (P > 0.01) indicating little genetic differentiation between Kazakh horses from the two countries.

Table 5. Pairwise F_{ST} values among populations.

Population*	AL	HE	TE	BL	HM	MN	TAL	ALM
AL								
HE	0.296**							
TE	-0.003	0.197**						
BL	0.114	0.145	0.001					
HM	0.093	0.161	0.006	-0.004				
MN	0.008	0.150	-0.040	-0.016	0.012			
TAL	0.119	0.126	0.014	-0.030	-0.014	-0.016		
ALM	0.082	0.143**	0.016	-0.014	0.033	0.006	-0.061	

Negative values equate to zero. **P < 0.01.

Phylogenetic analysis

The haplotypes of Kazak horses are grouped into six major lineages (Figure 2). Four of the six lineages (A, B, C, and D) corresponded to the four haplogroups (A, B, C, and D) shown in the MJ network (Figure 1). The other two lineages (E and F) corresponded to haplogroup E (Figure 1). Phylogenetic inference based on combined sequences supported the monophyly of most lineages (Figure 2). Five lineages (A to E) included haplotypes from horses in Kazakhstan and China. Haplotypes H9 and H43 were shared by horses in China and Kazakhstan. Lineage F only contained two haplotypes from China populations. The phylogenetic analyses showed that haplotypes from China and Kazakhstan populations were admixed. Five main clades (A-D and F, Figure 2) correspond to the haplogroups (A-D and F, Figure 1). Haplogroup E (Figure 1) contains not only clade E (Figure 2) but also some haplotypes that are not united in any of the above six clades, such as H7, H73, H26, and H64.

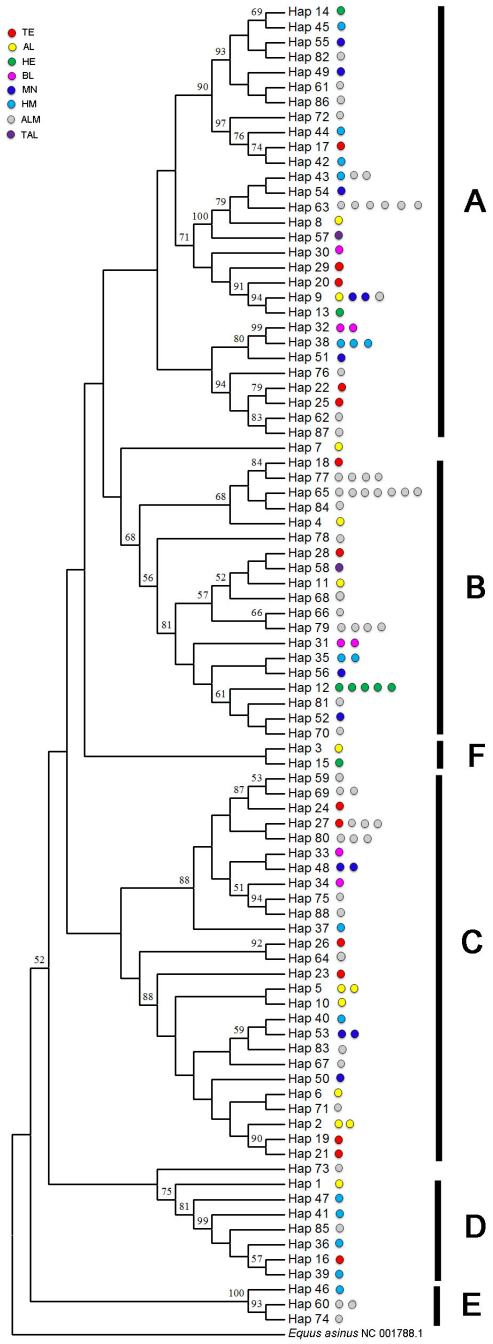

Figure 2. Consensus neighbor-joining tree of the 88 haplotypes. The tree is based upon 1000 bootstrap replicates. The reference *Equus asinus* sequence NC0017788.1 is an outgroup. Bootstrap values are shown as percentages above branches. The individuals with each haplotype are represented by colored circles indicating different populations. Different lineages (**A-F**) are indicated.

Relationships between Kazak and other horse breeds

Apart from Kazak breed, 25 horse breeds, Abaga, Akhal-teke, Arabian, Baise, Belgian, Caspian, Cleveland Bay, Clydesdale, Cheju, Debao, Exmoor, Friesian, Garrano, Haflinger, Jeju, Lusitano, Mongolian, Noriker, Przewalskii, SanHe, Shetland, Sorraia, Thoroughbred, Wuzhumuqin, and Yunnan, were used in the analyses. Within the Kazakh populations, 12 haplotypes were also found in individuals from a number of breeds, including Akhal-Teke, Cheju, Debao, Caspian, SanHe, Haflinger, Noriker, Arabian, and Sorraia (Figure 3). Some of the Kazakh horse haplotypes were either embedded or clustered with other breeds of horses (Figure 4).

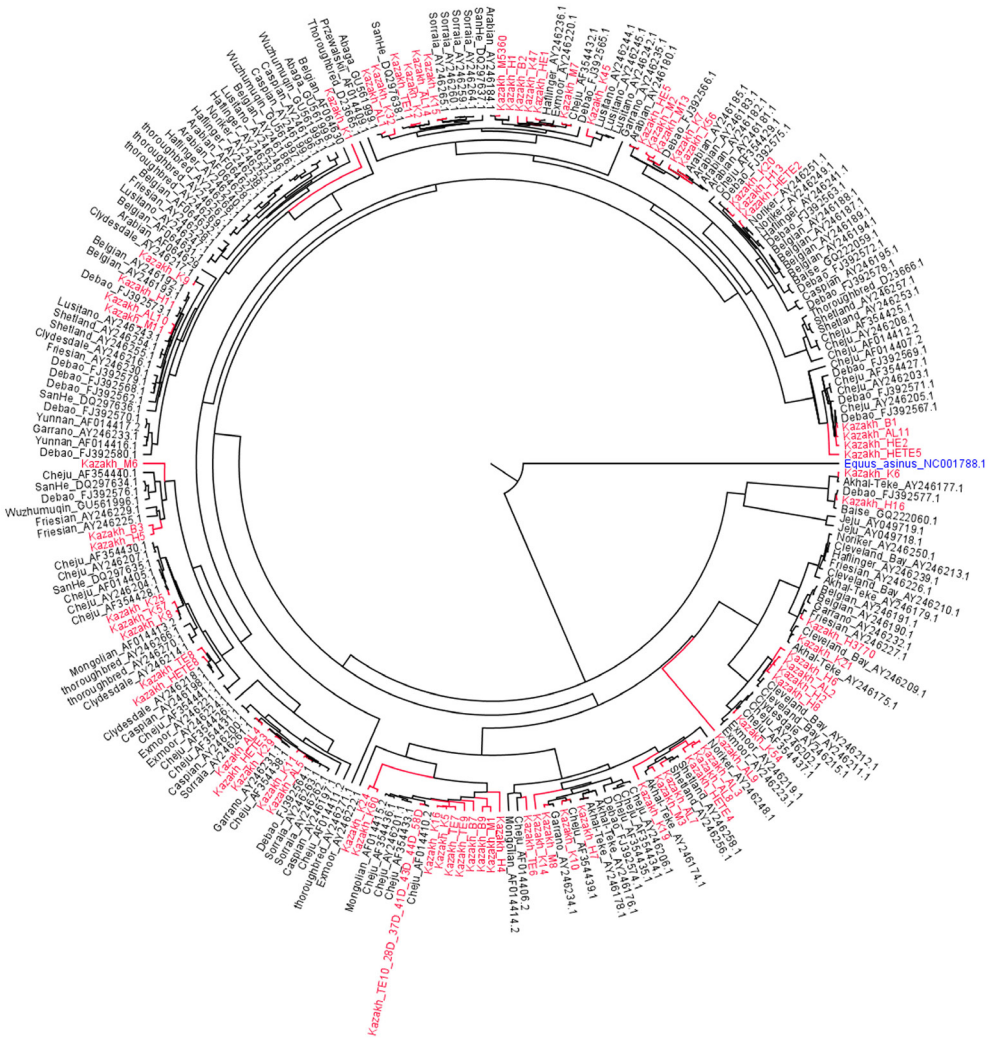

Figure 3. Phylogenetic relationships between the Kazakh horse breed and other horse breeds. BEAST phylogenetic tree was constructed with the D-loop region of 243 horse individuals including the Kazakh and 25 other breeds. Taxa names of Kazakh horses are indicated in red. *Equus asinus* (NC001788.1) was used as the outgroup.

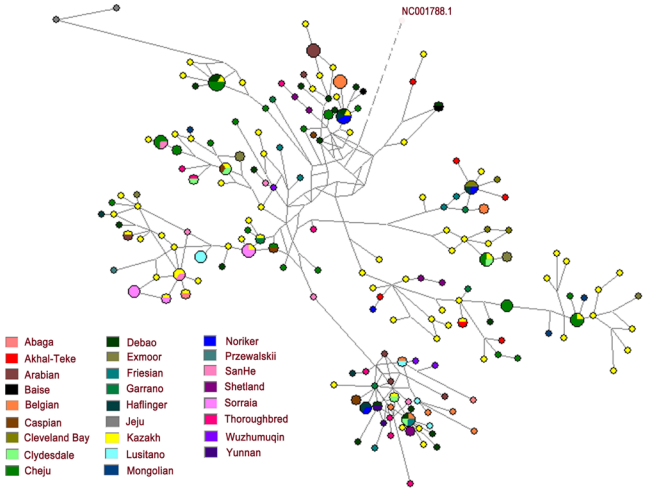

Figure 4. Median-joining network based on D-loop DNA sequences of 180 haplotypes representing 243 horse individuals from 26 horse breeds worldwide. Each population is indicated by a different color and the proportions of different populations within each haplotype are shown. NC001788.1 is the GenBank accession number of *Equus asinus* mtDNA.

DISCUSSION

Genetic diversity plays an important role in the survival and adaptability of a species (Frankham, 2005) Variation in the population's gene pool provides variable traits among the individuals of that population. These variable traits can be selected via natural selection, ultimately leading to an adaptive change in the population, allowing it to survive in a changed environment. If a population of a species has a highly diverse gene pool, individuals of that population will possess more variety of traits, thereby more traits for natural selection will act upon to select the fittest individuals to survive (Pullin, 2002). Genetic diversity is essential for a species to evolve. With very little genetic variation within a species, successful reproduction becomes increasingly difficult, and offspring are more likely to have issues resulting from inbreeding.

Some wild species, especially horse breeds, are subjected to small populations and inbreeding; thus, these species are facing the loss of genetic diversity. The Kazakh horse breed is an original local variety formed by long-term manipulation by people of all nationalities in Xinjiang. They are important to the local people, not only because they are a source of food (meat and milk), are an efficient tool in grazing or transport, and are used for horse breeding, but also they are of cultural and entertainment interest, playing an important role in the national culture, painting, music, dance, and horse racing.

Because of the lack of improvement, increase in consumption of horse meat, and excess hybridization with introduced horse breeds, the quality and quantity of Kazakh horses in China decreased dramatically, especially purebred Kazakh horses. It has been suggested that effective measures should be taken to protect and improve the Kazakh horses (Gemingguli, 2011). Accurate assessment of genetic diversity and structure is crucial to design and implement conservation strategies in natural and domesticated species. To date, there have been no studies on population structure and genetic diversity of Kazakh horses. Therefore, we

applied mitochondrial DNA data to investigate the genetic diversity of the Kazakh horse breed and its relationships with some other horse breeds.

A high level of genetic variability was observed in the Kazakh horses. Haplotype diversity among Kazakh horses was 0.9895, similar to that estimated in other horses, such as the Bardigiano horses (0.974), Italian heavy draught horses (0.991), Haflinger horses (0.926), Lipizzan horses (0.941), Maremmano horses (0.990), Murgese horses (0.964), Ventasso horses (0.900) (Bigi et al., 2014), Thoroughbred horses (1.00), Arab horses (0.96), Barb horses (1.00), and Andalusian horses (1.00) (Guastella et al., 2011). Among the haplotypes found in Kazakh horses, only three (H9, H27, and H43) are shared by horses in China and Kazakhstan. A higher haplotype diversity was found in Kazakh horse populations in China than in Kazakhstan. No congruence of haplogroup to a population's geographic origin was observed. Most of the haplogroups in the MJ network (Figure 1) contained haplotypes found both in China and Kazakhstan populations and no haplogroup included haplotypes from a single population. The six clades in the NJ tree largely corresponded with the haplogroups in the MJ network. However, the haplotypes were admixed and none of the phylogenetic clusters corresponded to populations of geographic areas or groups of the two countries. The pattern of broadly mixed haplotypes might suggest a recent history of hybridization or multiple origins of the Kazakh horse. *E. caballus* is thought to have multiple origins of domestication along with a strong bias for the recruitment of mares over stallions from the wild (Vilà et al., 2001; Jansen et al., 2002; Wallner et al., 2003; Cieslak et al., 2010).

Pairwise F_{ST} values between the populations and groups were estimated. Most populations show little differentiation with $F_{ST} < 0.05$ ($P > 0.01$) or moderate but not significant differentiation ($F_{ST} > 0.05$, $P > 0.01$). However, great differentiation was found between AL and HE populations, and HE and TE populations, both with $F_{ST} > 0.15$ ($P < 0.01$). Overall, little genetic differentiation was found between Kazakh horse groups from the two countries as the F_{ST} value between them was 0.02 ($P > 0.01$). AMOVA showed no subdivision between the two groups. None of the variance (-0.91%) is accounted for among the two groups, while a large fraction of variation (95.56%) was found within populations. Genetic distances also support that differentiation between Kazakh horses in China and Kazakhstan is very low. K-2-P distance between groups (1.0%) was not larger than distances within groups (1.0% each). Our results show that there is no clear pattern of differentiation among Kazakh horse populations from different geographic locations. This pattern indicates a close genetic relationship between the two horses from the two countries. This result might be because the two countries have frequent communication through the Silk Road in the past. In addition, long-term outcrossing and hybridization with introduced horses may be a factor. One study on maternal genetics of Kazakh horses revealed that the Kazakh horses have a rich maternal background, and the Chinese Kazakh horses and Kazakhstan Kazakh horses have close genetic relationships (Qi et al., 2014). Similar results are seen in other horse breeds, such as the Arabian horse (Khanshour and Cothran, 2013), the Cheju horse (Yang et al., 2002), and the Hispano-Breton heavy horse (Pérez-Gutiérrez et al., 2008). This can be explained by separate and geographically diverse populations participating in the domestication process or by domestic horses having multiple origins (Lister et al., 1998; Vilà et al., 2001; Jansen et al., 2002; Khanshour and Cothran, 2013).

In contrast with little differentiation in genetic variation and population structure, Kazakh horses in Kazakhstan and China have some obvious differences in their phenotype. After long periods of selective breeding, Kazakh horses in Kazakhstan have some better physical characteristics and performance than those in China. Individuals from Kazakhstan

are generally larger than those from China; for example, stallions are 500-600 kg and mares are 400-450 kg for Kazakhstan individuals, whereas stallions are 320-400 kg and mares are 260-350 kg for China horses. The average milk production of Kazakh horses in Kazakhstan is 12-15 L per day and only 3-8 L in China. The incongruence between genetics and phenotype can result from limited or unbalanced sampling. In the present study, we found that some haplotypes are unique in Kazakhstan populations. These genetic resources can be used for horse breeding or improvement of Kazakh horses in China in the future.

Genetic relationships between Kazakh and other breeds

Our results reveal that some of the Kazakh haplotypes are either shared or have close relationship with other kinds of horses. Within the Kazakh populations, 12 haplotypes were found in individuals from a number of breeds, including the Akhal-Teke, Cheju, Debao, Caspian, SanHe, Haflinger, Noriker, Arabian, and Sorraia (Figure 3), perhaps indicating that the Kazakh horse shares a common maternal ancestry with several different types of equines or there might have been gene flow among these horse breeds in the past. Further, a BEAST phylogenetic tree was constructed with the same sequence data. Phylogenetic analyses show that Kazakh horse mitochondrial sequences are widespread and distributed in many different clusters in the tree (Figure 3). Kazakh sequences cluster with many of the breeds, such as Akhal-Teke, Cheju, Debao, Caspian, SanHe, Haflinger, Noriker, Arabian, Sorraia Mongolian, and Thoroughbred, to form several clades. Similarly, little congruence of haplogroup distribution to breeds or geographic areas has been found as demonstrated in Cheju horses (Yang et al., 2002), Arabian horses (Khanshour and Cothran, 2013), German draught horses (Aberle et al., 2007), and Italian horses (Bigi et al., 2014). Achilli et al. (2012) analyzed 83 mitochondrial genomes from modern horses across Asia, Europe, the Middle East, and the Americas. Their data revealed 18 major haplogroups (A-R). Most of the 17 haplogroups identified in domestic breeds are spread over different geographic areas. This suggests that mitochondrial lineages are not powerful for identifying horse breeds because their diversification is thought to have occurred prior to domestication (Vilà et al., 2001; Jansen et al., 2002) or the existence of multiple maternal lines in the analyzed autochthonous breeds.

The shared haplotypes between Kazakh and other horse breeds and admixture between these breeds in phylogenetic analysis demonstrate past gene flow by hybridization with different horse breeds. The Kazakh horse breed is thought to have been influenced as early as the 5th century B.C. by many breeds, such as Mongolian, Karabair, Arabian, and Akhal-Teke. In the late 20th century, Kazakh horses were crossed with the Thoroughbred, Orlov Trotter, and Don (Dmitriez and Ernst, 1989) breeds. The development and success of the Kazakh horse breed is closely related to the nomadic transportation along the Silk Road and wars in these areas, as well as the vast lush grasslands and horse exchange between Xinjiang and Central Asia over the past several thousand years.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the International Cooperation Project of the Xinjiang Production and Construction Corps (Grant #2011BC007).

REFERENCES

Aberle KS, Hamann H, Drögemüller C and Distl O (2007). Phylogenetic relationships of German heavy draught horse breeds inferred from mitochondrial DNA D-loop variation. *J. Anim. Breed. Genet.* 124: 94-100. http://dx.doi.org/10.1111/j.1439-0388.2007.00636.x

Achilli A, Olivieri A, Soares P, Lancioni H, et al. (2012). Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. USA* 109: 2449-2454. http://dx.doi.org/10.1073/pnas.1111637109

Bandelt HJ, Forster P and Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16: 37-48. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026036

Bigi D, Perrotta G and Zambonelli P (2014). Genetic analysis of seven Italian horse breeds based on mitochondrial DNA D-loop variation. *Anim. Genet.* 45: 593-595. http://dx.doi.org/10.1111/age.12156

Bowling AT, Del Valle A and Bowling M (2000). A pedigree-based study of mitochondrial D-loop DNA sequence variation among Arabian horses. *Anim. Genet.* 31: 1-7. http://dx.doi.org/10.1046/j.1365-2052.2000.00558.x

Brown WM, George M, Jr. and Wilson AC (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76: 1967-1971. http://dx.doi.org/10.1073/pnas.76.4.1967

Cieslak M, Pruvost M, Benecke N, Hofreiter M, et al. (2010). Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One* 5: e15311. http://dx.doi.org/10.1371/journal.pone.0015311

DeAssis JB, DeLaat DM, Peixoto MGCD, Bergmann JAG, et al. (2009). Genetic diversity and population structure in Brazilian Mangalarga Marchador horses. *Genet. Mol. Res.* 8: 1519-1524. http://dx.doi.org/10.4238/vol8-4gmr647

Dmitriez NG and Ernst LK (1989). Animal genetic resources of the USSR. FAO Animal Production and Health Paper 65. Food and Agriculture Organization of the United Nations, Rome.

Drummond AJ, Suchard MA, Xie D and Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29: 1969-1973. http://dx.doi.org/10.1093/molbev/mss075

Excoffier L and Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10: 564-567. http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x

Frankham R (2005). Genetics and extinction. *Biol. Conserv.* 126: 131-140. http://dx.doi.org/10.1016/j.biocon.2005.05.002

Gemingguli M (2011). A comparative analysis of production performance of the Kazakh horses. *China Herbivores Sci.* 31: 69-71 (In Chinese).

Gizaw S, Komen H, Windig JJ, Hanotte O, et al. (2008). Conservation priorities for Ethiopian sheep breeds combining threat status, breed merits and contributions to genetic diversity. *Genet. Sel. Evol.* 40: 433-447.

Guastella AM, Zuccaro A, Criscione A, Marletta D, et al. (2011). Genetic analysis of Sicilian autochthonous horse breeds using nuclear and mitochondrial DNA markers. *J. Hered.* 102: 753-758. http://dx.doi.org/10.1093/jhered/esr091

Hall SJG and Bradley DG (1995). Conserving livestock breed biodiversity. *Trends Ecol. Evol.* 10: 267-270. http://dx.doi.org/10.1016/0169-5347(95)90005-5

Hartl D and Clark A (1997). Principles of population genetics. Sinauer Associates Inc., Sunderland.

Hutchison CA, Newbold JE, Potter SS and Edgell MH (1974). Maternal inheritance of mammalian mitochondrial DNA. *Nature* 251: 536-538. http://dx.doi.org/10.1038/251536a0

Jansen T, Forster P, Levine MA, Oelke H, et al. (2002). Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl. Acad. Sci. USA* 99: 10905-10910. http://dx.doi.org/10.1073/pnas.152330099

Khanshour AM and Cothran EG (2013). Maternal phylogenetic relationships and genetic variation among Arabian horse populations using whole mitochondrial DNA D-loop sequencing. *BMC Genet.* 14: 83. http://dx.doi.org/10.1186/1471-2156-14-83

Librado P and Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452. http://dx.doi.org/10.1093/bioinformatics/btp187

Lippold S, Matzke NJ, Reissmann M and Hofreiter M (2011). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol. Biol.* 11: 328. http://dx.doi.org/10.1186/1471-2148-11-328

Lister AM, Kadwell M, Kaagan LM and Stanley HF (1998). Ancient and modern DNA in a study of horse domestication. *Anc. Biomol.* 2: 267-280.

Mirol PM, Peral García P, Vega-Pla JL and Dulout FN (2002). Phylogenetic relationships of Argentinean Creole horses and other South American and Spanish breeds inferred from mitochondrial DNA sequences. *Anim. Genet.* 33: 356-363. http://dx.doi.org/10.1046/j.1365-2052.2002.00884.x

Ning T, Xiao H, Li J, Hua S, et al. (2010). Adaptive evolution of the mitochondrial ND6 gene in the domestic horse. *Genet. Mol. Res.* 9: 144-150. http://dx.doi.org/10.4238/vol9-1gmr705

Ocokoljic M, Vilotic D and Sijacic-Nikolic M (2013). Population genetic characteristics of horse chestnut in Serbia. *Arch. Biol. Sci.* 65: 1-7. http://dx.doi.org/10.2298/ABS1301001O

Pérez-Gutiérrez LM, De la Peña A and Arana P (2008). Genetic analysis of the Hispano-Breton heavy horse. *Anim. Genet.* 39: 506-514. http://dx.doi.org/10.1111/j.1365-2052.2008.01762.x

Plante Y, Vega-Pla JL, Lucas Z, Colling D, et al. (2007). Genetic diversity in a feral horse population from Sable Island, Canada. *J. Hered.* 98: 594-602. http://dx.doi.org/10.1093/jhered/esm064

Prystupa JM, Hind P, Cothran EG and Plante Y (2012). Maternal lineages in native Canadian equine populations and their relationship to the Nordic and Mountain and Moorland pony breeds. *J. Hered.* 103: 380-390. http://dx.doi.org/10.1093/jhered/ess003

Pullin AS (2002). Conservation biology. 1st edn. Cambridge University Press, Cambridge.

Qi A, Wumaierjiang A, Iskhan K, Cheng L, et al. (2014). Maternal genetics of the Kazakh horses in China and Kazakhstan. *Heilongjiang Xumu Shouyi* 7: 31-34 (In Chinese).

Takasu M, Ishihara N, Tozaki T, Kakoi H, et al. (2014). Genetic diversity of maternal lineage in the endangered Kiso horse based on polymorphism of the mitochondrial DNA D-loop region. *J. Vet. Med. Sci.* 76: 1451-1456. http://dx.doi.org/10.1292/jvms.14-0231

Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729. http://dx.doi.org/10.1093/molbev/mst197

Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680. http://dx.doi.org/10.1093/nar/22.22.4673

Tiemuerbai Y (2014). On the protection of the Kazakh horse breeds. *Alrai Newspaper*, 15th, April 1-2 (In Chinese).

Vilà C, Leonard JA, Gotherstrom A, Marklund S, et al. (2001). Widespread origins of domestic horse lineages. *Science* 291: 474-477. http://dx.doi.org/10.1126/science.291.5503.474

Wallner B, Brem G, Müller M and Achmann R (2003). Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. *Anim. Genet.* 34: 453-456. http://dx.doi.org/10.1046/j.0268-9146.2003.01044.x

Winton CL, Hegarty MJ, McMahon R, Slavov GT, et al. (2013). Genetic diversity and phylogenetic analysis of native mountain ponies of Britain and Ireland reveals a novel rare population. *Ecol. Evol.* 3: 934-947. http://dx.doi.org/10.1002/ece3.507

Wright S (1978). Evolution and the genetics of population, variability within and among natural populations. The University of Chicago, Chicago.

Yang YH, Kim KI, Cothran EG and Flannery AR (2002). Genetic diversity of Cheju horses (*Equus caballus*) determined by using mitochondrial DNA D-loop polymorphism. *Biochem. Genet.* 40: 175-186. http://dx.doi.org/10.1023/A:1015884125507

Zhao TZ (1991). The Kazakh horses. *Xinjiang Agric. Sci.* 5: 222-223 (In Chinese).

Supplementary material

Table S1. D-loop sequences downloaded from Genbank and accession numbers.