Current genetic conservation of Chinese indigenous horses revealed with Y-chromosomal and mitochondrial DNA polymorphisms

Shuqin Liu,[1,2,3,†] Chunzheng Fu,[1,2,†] Yunzhou Yang,[1,2] Yuanyuan Zhang,[2] Hongying Ma,[1,2] Zhiyao Xiong,[1,2] Yao Ling,[1,2,4,5] and Chunjiang Zhao [1,2,4,5,6,*]

[1]Equine Center, China Agricultural University, Beijing, China
[2]College of Animal Science and Technology, China Agricultural University, Beijing, China
[3]College of Animal Science and Technology, Qingdao Agricultural University, Shandong, China
[4]Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, China
[5]Beijing Key Laboratory of Animal Genetic Improvement, Beijing, China
[6]National Engineering Laboratory for Animal Breeding. Beijing, China
[†]These authors contributed equally to this work.

*Corresponding author: cjzhao@cau.edu.cn

Abstract

To investigate the genetic diversity of Chinese indigenous horses and determine the genetic status of extant horse breeds, novel Y chromosomal microsatellite markers and known Y chromosomal SNPs and mtDNA loop sequences, were employed to study the genetic diversity levels of 13 Chinese indigenous horse populations and four introduced breeds. Sixteen Y-chromosomal microsatellite markers, including seven newly identified loci, were used in the genotyping. The results showed that 4 out of the 16 loci were highly polymorphic in Chinese indigenous horse populations, in which the polymorphisms of 3 loci, ECAYP12, ECAYP13, and ECAYCAU3, were first reported in the present study. The polymorphic Y chromosomal microsatellite markers result in 19 haplotypes in the studied horses and formed 24 paternal lines when merged with the 14 Y chromosomal SNPs reported previously. The haplotypes CHT18 and SS24 harboring *AMELY* gene mutation were the ancestral haplotypes, and other haplotypes were derived from them by one or more mutation steps. The horse populations in mountainous and remote areas of southwestern China have the most ancient paternal lines, which suggests that ancient paternal lines preserved in local populations attributed to less human interventions. Our results also showed that the northern local breeds had higher mtDNA diversity than the southern ones in China. The frequency of haplogroup B, F, and G of mtDNA in Chinese indigenous horses has declined in recent years, and some breeds are in endangered status mainly due to small population sizes. Urgent actions should be taken to conserve the genetic diversity of the indigenous horse populations, especially the rare paternal lines. Our findings help to elucidate the genetic diversity and evolutionary history of Chinese domestic horses, which will facilitate the conservation of the indigenous horses in the future.

Keywords: horse; Y chromosome; microsatellite; SNP; mtDNA

Introduction

China had the largest number of horses in the world in the 1970s according to the UN Food and Agriculture Organization (FAO, http://www.fao.org/home/en/), but the horse population decreased by over 50% between 1977 and 2017 (FAO 2017; National Bureau of Statistics, 2017). Although the total population size of Chinese horses is still large, there were 11 horse breeds with small population sizes, in each of which breeding stallions were less than 20. The endangered and critical Chinese local breeds account for up to 26% of the total horse populations in China (Scherf 2000; Chang 2011).

The genetic diversity in local livestock species was caused by mutation, selective breeding, isolation, genetic drift, and migration. The diversity provides valuable resources for breeding programs and enables livestock populations to adapt as environmental conditions change (Rischkowsky and Pilling 2007). Studies of mtDNA revealed that extant horses have the highest mtDNA diversity among domestic species (Vila *et al.* 2001; Jansen *et al.* 2002; Cieslak *et al.* 2010), while notably low-genetic variation was detected in horse Y chromosomes (Wallner *et al.* 2003, 2004; Lindgren *et al.* 2004; Lippold *et al.* 2011; Rafeie *et al.* 2011; Brandariz-Fontes *et al.* 2013; Wutke *et al.* 2018). Autosomal diversity, revealed with polymorphism of autosomal markers, which is partitioned among horse breeds and may reflect breeding activities (Vila *et al.* 2001), showed that the genetic diversity in horses severely declined over the past 400 years (Fages *et al.* 2019). Nucleotide diversity of horse Y chromosome also had decreased steadily for about 2000 years but dropped rapidly to present-day levels only after 850–1350 CE (Fages *et al.* 2019).

In some studies, no polymorphism of the horse Y chromosome was identified by screening the sequence of chromosomes with sequencing or genotyping microsatellite DNA markers (Wallner et al. 2003; Lindgren et al. 2004; Rafeie et al. 2011; Brandariz-Fontes et al. 2013). The first DNA variation of the horse Y chromosome, a variant of Y chromosomal microsatellite (also called Y chromosomal short tandem repeat, Y-STR), was found in Chinese indigenous populations in 2010 (Ling et al. 2010). In the study conducted by Wallner et al. (2013), only four polymorphic loci, including three SNPs and a single base deletion, were found in the 181-kb sequenced DNA region of the horse Y chromosome, resulting in 6 haplotypes, but no Y-chromosomal microsatellite variation was detected. Following the above results, Han et al. studied the paternal lines of Chinese indigenous horses and detected the same three Y-chromosome DNA variants reported by Wallner et al. (2013), but they formed five haplotypes that were different from those found in European horses (Han et al. 2015). Based on the extended male-specific Y (MSY) sequence scope, a number of Y chromosomal SNPs (Y-SNPs) were identified in modern horses, and it was also found that Asian horses have much higher diversity than European breeds (Wallner et al. 2017; Felkel et al. 2018, 2019). Kreutzmann et al. (2014) isolated 8 Y-chromosomal microsatellites and then screened 104 male horses from 41 breeds, most of which were European, and only two haplotypes were identified. However, Y-chromosomal diversity might be underestimated on the sequence level due to ascertainment bias (Lenstra et al. 2012). Further study including more Y-chromosomal sequences and horse breeds, ideally indigenous breeds, is needed to obtain more detailed information about horse paternal diversity and evolution.

In this study, we investigated the genetic diversity of 13 Chinese indigenous horse breeds with Y-STRs and Y-SNPs as well as mtDNA to reveal current situations of the genetic preservation of the indigenous horse breeds, in order to provide valuable information for the genetic conservation of Chinese indigenous horses.

Materials and methods
Sample collection and DNA extraction
A total of 801 blood samples from 13 indigenous breeds in China and four introduced breeds and a donkey breed were involved in the present study, and the detailed sample information was provided in Table 1. Most of the indigenous horse samples were from our previous studies (Yang et al. 2017). The indigenous horses were sampled in their distributed areas (Table 1 and Supplementary Figure S1). To avoid close blood relationship, we kept at least 500 meters between the sampling sites. But the small male sample sizes (<5 individuals) of the three breeds, IMG, NQ, and AR, may cause possible bias on the estimation of the genetic diversity. A donkey, sampled in Shandong Province, was used as an outgroup. China Stud Book provided the samples of the introduced horses and their original geographic information. The samples were randomly selected from the introduced individuals without close kinship. All samples were collected according to the rules of the ethics committee of China Agricultural University. The samples for mtDNA studies were from both genders, and only male samples were used for Y-STR and Y-SNP analysis.

According to horses' geographical distribution, appearance, and historical information, domestic horse breeds in China were divided into five groups: Mongolia Group (MG), Kazak Group (KZKG), Hequ Group (HQG), Tibet Group (TIG), and Southwest

Group (SWG). MG, KZKG, and HQG are distributed in North China while TIG and SWG consist of southern horses (Table 1).

Genomic DNA was isolated from samples using the phenol-chloroform method and preserved at −20°C. DNA was quantified using a NanoDrop 2000 (USA). DNA samples with A260/A280 values between 1.8 and 2.1 were adjusted to a concentration between 20 and 30 ng/μl. The quality of DNA samples was further evaluated and confirmed by agarose gel electrophoresis. Furthermore, all male samples were verified by PCR with the Y chromosome-specific primers ECAY2B17 (Wallner et al. 2003).

Y chromosomal microsatellite genotyping
Sixteen Y chromosomal microsatellite markers were used to analyze the genetic diversity of the horses (Supplementary Table S1). Nine of the 16 loci (Loci ECAYA16, ECAYE2, ECAYE3, ECAYNO1, ECAYNO2, ECAYNO3, ECAYNO4, ECAYP12, and ECAYP13) were reported in previous studies (Wallner et al. 2003; Kreutzmann et al. 2014), and the published primers were used in this study, except that the primers for ECAYP12 were redesigned. The other seven novel microsatellite loci, ECAYCAU1, ECAYCAU2, ECAYCAU3, ECAYCAU4, ECAYCAU5, ECAYCAU6, and ECAYCAU7, were identified and first reported here (Supplementary Table S1). The novel microsatellite markers were detected by screening the sequences of horse Y-chromosomal BAC retrieved from GenBank with SSRHunter software. Their Y chromosome specificity was confirmed by PCR amplification of genomic DNA from stallions, female horses, and blank control samples (Supplementary Figure S2). The primers for the new makers were designed with Primer Premier 5.0 (Premier Biosoft International). The forward primers of the polymorphic loci were 5'-labeled with fluorescent dyes (TAMRA, TET, 6-FAM, or HEX), used for genescanning analysis. Unlabeled primers were used in other experiments in the present study (such as PAGE-silver staining mentioned below). All of the primers were synthesized by Sangon Biotech (Shanghai, China).

PCR amplifications were performed in a 20 μl reaction consisting of 25 ng DNA, 0.5 μM each primer, 10 μl 2 X Taq PCR MasterMix (Tiangen KT201, Beijing, China) with 0.1 U Taq Polymerase/μl, 3 mM Mg^{2+}, 500 μM dNTPs and 8 μl ddH₂O. All amplifications included an initial denaturing step of 10 min at 95°C followed by 35 cycles of 30 s at 95°C, 30 s at an optimized annealing temperature, as shown in Supplementary Table S1, and 30 s at 72°C, ended with a final extension of 10 min at 72°C. Amplifications were carried out in ABI thermal cyclers (ABI, USA), and 2.5 μl PCR products were visualized by electrophoresis in 2% agarose gel to ensure amplification success. The samples were screened for variants with PAGE in the initial step, in which 12% polyacrylamide gel electrophoresis was pre-run at 150 V for 30 min firstly, and then 2.5 μl PCR products were added, and the electrophoresis was conducted at 120 V for 8–9 h. Gels were silver-stained according to the method of Byun et al. (2009). PCR products harboring DNA variants were further tested by genescanning with capillary electrophoresis using an ABI 3730XL DNA Genetic Analyzer (USA) (genescanning method). Peak Scanner Software v1.0 was used to analyze the microsatellite data.

For further confirmation, the PCR products carrying different variants were also ligated into the PMD 18-T vector (TaKaRa, Japan). Following the transformation of the ligation products, plasmids were harvested from the resulting bacterial colonies and sent to BGI (Beijing, China) for sequencing.

Table 1 Equine samples analyzed in the present study

Horse Group (abbr.)		Breed name (abbr.)	Location (county, province)	mtDNA[a]	Y-SNP[b]	Y-STR[c]
North China	Mongolia Group (MG)	Mongolia (IMG)	Chifeng, Inner Mongolia	49	7	2
		E lunchun (ELC)	E lunchun, Inner Mongolia	29	10	5
	Kazak Group (KZKG)	Yanqi (YQ)	Yanqi, Xinjiang	43	8	6
		Kazak (KZK)	Yili, Xinjiang	47	12	7
	Hequ Group (HQG)	Chakouyi (CK)	Tianzhu, Gansu	23	9	9
South China	Tibet Group (TIG)	Naqu (NQU)	Naqu, Tibet	27	7	9
	Southwest Group (SWG)	Baise (BS)	Baise, Guangxi	19	9	10
		Debao pony (DBP)	Debao, Guangxi	108	37	80
		Lijiang (LJ)	Lijiang, Yunnan	91	34	36
		Tengchong (TC)	Tengchong, Yunnan	37	25	33
		Guizhou (GZ)	Bijie, Guizhou	75	13	12
		Jinjiang (JJ)	Jinjiang, Fujian	21	10	9
		Ningqiang (NQ)	Ningqiang, Shanxi	40	6	2
Foreign countries	AR	Arabian horse (AR)	Xinjiang (introduced from Russia)	89	15	3
	TB	Thoroughbred (TB)	Beijing (introduced from Ireland)	56	31	16
	AT	Akhal Teke horse (AT)	Xinjiang (introduced from Russia)	47	13	17
	WB	Warmblood horse (WB)	Shandong (introduced from Germany)	0	7	12
AS		Donkey (*Asinus*, AS)	Liaochen, Shandong	0	1	1
Total number		21	21	801	254	269

[a] Indicates data of mtDNA cited from Yang *et al.* (2017), except Foreign horses, which were newly tested in the present study;
[b] The Y-SNP data were mainly from our previous study (Liu *et al.* 2020).
[c] 165 of the samples yielded both the Y-STR and Y-SNP data.

Y chromosomal SNP genotyping

The genotyping data of 14 Y chromosomal SNP sites were mainly from our previous study (Liu *et al.* 2020), while some additional samples of male horses and one donkey sample were also tested in the present study following the methods reported previously (Liu *et al.* 2020; Table 1). Eventually, 253 indigenous horses and 66 introduced horses and one donkey were successfully genotyped at all the 14 Y chromosomal sites (Table 1), which were amplified with nine pairs of primers, including Pre34, H15, H2, IN2-2, N23-5, M23-4, Y288, Y869, and Y997 (Liu *et al.* 2020).

Statistical analysis

Nucleotide diversity (π) and haplotype diversity (h) within each population were calculated by DnaSP5 software (Librado and Rozas 2009). And the π and h of the merged populations (the merged three populations include horses of North China, South China, and the introduced horses from foreign countries) were analyzed with the pegas package in the R project (Paradis 2010). The default settings were used for the software. The diversities of merged populations were statistically analyzed with the T test by hand. In order to compare the number of Y-SNP and Y-STR haplotypes among breeds with different sample sizes, the sample sizes were normalized by resampling the populations with 1000 replication for an equal number of samples, applying with an in-house developed program (available on Github, https://github.com/yyzhangcau/horseGeneticDiversity-Zhao). The mtDNA haplotypes were grouped into one of seven major haplogroups, as described by Jansen *et al.* while some haplotypes were classed as "other" groups, which did not belong to the 19 known haplotypes (Jansen *et al.* 2002; Cieslak *et al.* 2010). The frequencies of the mtDNA-loop haplogroups in the three merged horse populations were analyzed with Chi-square test of SPSS to determine the different level of their haplogroup frequencies. F-statistics (Fst) based on the Y-STRs was calculated by Arlequin 3 with haploids of the makers (Excoffier and Lischer 2010). To analyze the data

with Arlequin 3, we used following settings: population comparisons, computing pairwise FST, and 1000 permutations. The level of significance was defined at $P < 0.05$. The values determined by the number of different alleles (FST-like) or sum of squared differences (RST-like) were calculated separately. For the genetic diversity of Y-STRs, the number of alleles, the number of effective alleles (Ne) per locus, polymorphism information content (PIC), and Shannon index were calculated. The number of alleles per locus was obtained by directly counting. Ne, PIC, and Shannon index were calculated using the GENEPOP package version 1.32 (Yeh *et al.* 1997). Nei's distance (DA) was determined by the unweighted pair-group method with arithmetic mean (UPGMA), and a dendrogram was built employing POPTree (Nei *et al.* 1983), and tree robustness was assessed by 1000 bootstrap replicates. To analyze the population relationship among Y haplotypes with the genotyping data of Y-STRs or the combining data from the Y-SNPs and Y-STRs, we reconstructed median-joining networks using NETWORK 5.0 (Bandelt *et al.* 1999), and default settings were chosen.

Data availability

SNP variant data are available from EVA (Project: PRJEB35322). mtDNA data from the previous study can be obtained at https://doi.org/10.1111/asj.12583. Supplementary Figure S1 shows the geographic location of the studied Chinese indigenous horse breeds. In the figure, ELC indicates E lunchun; IMG, Inner Mongolia; YQ, Yanqi; KZK, Kazak; CK, Chakouyi; NQU, Naqu; LJ, Lijiang; BS, Baise; DBP, Debao pony; TC, Tengchong; GZ, Guizhou; JJ, Jinjiang; NQ, Ningqiang. Supplementary Figure S2 displays Y-specificity of the newly identified microsatellite markers (ECAYCAU1-7), in which M indicates the DNA standards (DNA marker 2000); m, male horse; f, female horse; c, blank control. Supplementary Figure S3 illustrates the alleles of the polymorphic loci shown with genescanning peaks. In Supplementary Figure S3, a1–a4 indicates the four alleles of ECAYA16, 155, 153,

151 bp, and no peak (null) at the locus; b1–b4, the four allele patterns of ECAYP12, including b1 (208 bp), b2 (204 and 208 bp), b3 (200 and 208 bp), and b4 (200, 204, and 208 bp); c1–c3, the three allele patterns of ECAYP13, which are c1 (106 bp), c2 (115 bp), and c3 (115 and 118 bp), respectively; d1–d3, the three alleles of ECAYCAU3 consisting of 128, 130, and 132 bp. Supplementary Figure S4 shows the sequencing results of the PCR products harboring variants of the four polymorphic Y-STRs. Supplementary Figure S5 shows a median-joining network based on the polymorphisms of the 16 Y-STRs from the 269 equid samples. In Supplementary Figure S5, Y-STR haplotypes are represented by circles with areas proportional to the number of individuals harboring the haplotype. Colors indicate the breeds. The number of mutations was shown in branches. IN indicates the introduced horses. Supplementary Figure S6 displays the Matrix of pairwise Fst between different horse breeds based on Y-STR data. In the figure, ELC indicates E lunchun; IMG, Inner Mongolia; YQ, Yanqi; KZK, Kazak; CK, Chakouyi; NQU, Naqu; LJ, Lijiang; BS, Baise; DBP, Debao pony; TC, Tengchong; GZ, Guizhou; JJ, Jinjiang; NQ, Ningqiang; AR, Arabian horse; TB, Thoroughbreds; AT, Akhal Teke; WB, Warm Blood.

Supplementary Table S1 shows the information on the horse Y chromosomal microsatellite loci and primers used in this study. Supplementary Table S2 lists the Y-STR haplotypes harbored by the studied equids. Supplementary Table S3 shows the nucleotide and haplotype diversity of mtDNA and Y-chromosomal SNPs in the horses of North China, South China, and the introduced horses. Supplementary Table S4 shows pairwise Fst and Rst values between the studied horse breeds calculated with Y-STR data. Supplementary Table S5 lists the Y-SNP-STR haplotypes harbored by the studied equids. Supplementary Table S6 shows the frequencies of the mtDNA-loop haplogroups in the studied horse populations. Supplemental Material available at figshare: https://doi.org/10.25387/g3.11322890.

Results
The variation of Y chromosomal microsatellites

The genotyping results showed four loci (ECAYA16, ECAYP12, ECAYP13, and ECAYCAU3) out of the 16 Y chromosomal microsatellites polymorphic in the 268 studied male horses (Supplementary Figure S3), and the sequences of the four loci are shown in Supplementary Figure S4. Loci ECAYA16 and ECAYP12 both have four alleles, and 155 and 208 bp are their prominent alleles, respectively. There are three alleles detected at each of the loci ECAYP13 and ECAYCAU3, and their prominent alleles are 115 and 128 bp, respectively. Detailed information for the loci is shown in Table 2. However, loci ECAYP12 and ECAYP13 of some samples in this study exhibited double or triple basecalls, which resulted from duplications in the male-specific Y chromosome (Kayser et al. 2000). Two or three similar copies of one fragment with different alleles were made to constitute a unique haplotype used to calculate genetic diversity and reconstruction of phylogenetic structure (Supplementary Table S2).

As shown in Table 2, there are 3 or 4 alleles for each of the polymorphic loci, and ECAYP12 possesses the greatest number of effective alleles (Ne), Shannon index, and the PIC (Tables 2 and 3). The variants of the four polymorphic markers were only detected in Chinese indigenous horses, and all four introduced horse breeds were monomorphic (Table 3). IMG, ELC, and NQ did not display variations in all the Y-STR loci, most likely due to the small sample sizes or the massive impact of modern breeding.

Furthermore, most horse breeds of SWG group showed high diversity at the Y-STR and SNP loci (Table 3). GZ showed the highest variations at ECAYA12 locus (PIC = 0.62). Overall, the southern breeds as a whole harbor a higher level of Y-chromosomal diversity than the breeds of North China and the foreign breeds, which is contrary to results from mtDNA data (P < 0.05) (Table 3 and Supplementary Table S3).

The polymorphisms of the Y-STRs resulted in 19 haplotypes (CHT1 to CHT19) in horses and one (CHT20) in a donkey (Supplementary Figure S5 and Supplementary Table S2). A median-joining network based on the genotyping results of the 16 Y-STRs was constructed. CHT1 and CHT7 are the two most predominant haplotypes, which contain 134 and 76 horses, respectively. All the introduced horses, including Arabian horses, Thoroughbreds, Warmblood, and Akhal Teke horses are distributed in the haplotype CHT1. Each of other haplotypes is harbored by less than 13 individuals. Seven haplotypes (accounting for 35% of the Y-STR haplotypes) were unique and each of them is carried by only one individual. In the studied breeds, Debao and Guizhou horse breeds harbored 10 and 7 haplotypes (4.8 and 5.2 haplotypes after adjusted for the difference in sample sizes), respectively, and had more abundant Y-STR diversity than other breeds. CHT20 is the only haplotype carried by a donkey, which is the most ancestral haplotype. CHT18, harbored by a Guizhou horse, roots from CHT20 directly and is the most ancient haplotype in the horses. The two most common haplotypes, CHT1 and CHT7, are relatively young patrilines. Based on the network analysis, the haplotypes of the studied horses formed four haplogroups (CHTG1 to CHTG4). CHTG1, containing southern horse breeds, is the relatively ancient haplogroup among the four haplogroups. All other haplogroups are derived from CHTG1 by one or more steps of mutations.

The Fst calculated with the STR data indicated that the genetic background differences mainly exist between breeds with far geographic distance. There are the smallest Fst values between the introduced breeds or the breeds from the SWG region (DBP, LJ, TC, and GZ), and significant genetic differences were observed between the indigenous breeds and the introduced breeds (Supplementary Figure S6, Supplementary Table S4). Based on the Fst results, cluster analysis was performed using regional population data. Figure 1 shows that the studied horse breeds were divided into two distinct groups according to the Y-STR polymorphic information. One group consists of Chinese indigenous groups, and another only contains the four introduced horse breeds. In Chinese horses, SWG horses harbor more distant paternal lines than others, while MG, HQG, and TIG cluster into one group and KZKG forms another cluster.

Paternal Y-SNP-STR network

In our previous work, a total of 14 horse Y-SNPs were successfully genotyped (Liu et al. 2020). A mutation was detected in the highly conserved gene Amelogenin (Y-linked; AMELY). Nine samples from the six southern horse populations, including DBP, BS, LJ, and GZ, harbor the C allele (AB091794: 2614 bp), while the majority of the studied individuals carry the T allele at the locus.

Integrating STR and SNP markers of the Y chromosome, 25 haplotypes resulted in the studied equids (164 horses and one donkey) (Figure 2, Supplementary Table S5). Compared with the northern horses, more haplotypes were detected in the southern populations. Ten out of the 25 haplotypes are unique, and each of them is harbored by only one individual. Seven unique haplotypes were detected in the southern populations. SS22 (STR and SNP haplotype 22, abbreviated as SS22), SS23, and SS24, all of

Table 2 Observed allele repeated motif, sizes, allele frequencies and genetic parameters of the four polymorphic Y-specific microsatellite loci in the studied equids

Locus	Repeated motif	Allele (bp)	Allele frequency (%)	Ne	Shannon
ECAYA16	(TG)3TAT(GT)16	151	5.22	1.20	0.40
	(TG) 3TAT(GT)17	153	0.75		
	(TG)3TAT(GT)18	155	91.04		
	null	0	2.99		
ECAYP12	(GATA)14	208	53.36	2.41	0.98
	(GATA)12, (GATA)14	200/208	37.69		
	(GATA)13, (GATA)14	204/208	8.21		
	(GATA)12, (GATA)13, (GATA)14	200/204/208	0.75		
ECAYP13	(TAA)5T(TAA)7	106	0.37	1.17	0.30
	(TAA)5T (TAA)10	115	92.91		
	(TAA)5T(TAA)10, (TAA)5T(TAA)11	115/118	6.72		
ECAYCAU3	(TG) 19	128	94.40	1.13	0.26
	(TG) 20	130	0.75		
	(TG) 21	132	4.85		

which carry C alleles of *AMELY*, were the closest haplotypes to the donkey. SS2 and SS5 contain the most introduced and indigenous horses.

Diversity of mtDNA and the distribution of the haplogroups

The mtDNA nucleotide and haplotype diversity of the horse populations of North China, South China, and the introduced horses were calculated with the data of the mitochondrial D-loop hypervariable segment 1. And the results showed that the northern horses had significantly higher maternal diversity than the southern and introduced groups (Supplementary Table S3). The frequencies of the mtDNA haplogroups in the studied horse groups indicated that both the introduced horses and the Chinese indigenous horses harbor all of the seven mtDNA haplogroups (Supplementary Table S6). However, the northern Chinese horses carry a higher frequency of B haplogroup than the southern group or the introduced horses ($P < 0.01$), while F haplogroup, which was regarded as a maternal line of oriental-origin, also shows the highest frequency in the northern horses among the three groups ($P < 0.05$). In all of the seven haplogroups, G is the rarest. No individual harboring G haplogroup was identified in the studied introduced horses. Low frequencies of the rare haplogroup were detected in only six populations out of the 13 Chinese indigenous breeds (Supplementary Table S6).

Discussion
High patrilineal diversity of indigenous horses in China

Previous studies showed notably low diversity of horse Y-chromosomal sequence (Wallner *et al.* 2003, 2004; Lindgren *et al.* 2004; Wallner *et al.* 2004; Lippold *et al.* 2011; Rafeie *et al.* 2011; Brandariz-Fontes *et al.* 2013), while Asian horses had much higher diversity than European breeds (Felkel *et al.* 2018), and European horses were influenced by oriental horses (Wallner *et al.* 2017; Fages *et al.* 2019). In the past several hundred years, horses in Europe have been subjected to intensive breeding activities, in which severe selection was imposed on the males, and only the superior male horses had the chance to breed (Hendricks 2007; Hamann and Distl 2008). Introduced stallions, such as the stallions of Arabian and Thoroughbred horses, have been widely used to improve the sport traits of local horses in European countries. As a result, the indigenous paternal lines were replaced by the foreign ones in many European breeds. Compared with most

horse breeds that have experienced intensive artificial selection and genetic introgression (Hendricks 2007), Chinese indigenous horse breeds have not been confronted with high selection pressure and were bred under various ecological conditions and demonstrated a high level of morphological and genetic diversity (Lei *et al.* 2009; Jiang *et al.* 2011; Ling *et al.* 2011; Xu *et al.* 2012; Yue *et al.* 2012).

In this study, the indigenous horses showed relatively higher Y-chromosomal genetic diversity than the introduced horse breeds. A few southern horses harbor C allele of *AMELY* gene, which were only observed previously in ancient and *Przewalski's* horses (Lindgren *et al.* 2004; Lippold *et al.* 2011; Brandariz-Fontes *et al.* 2013). We also identified four highly polymorphic Y-chromosomal microsatellite loci by screening Chinese indigenous horses. Variants of loci ECAYP12, ECAYP13, and ECAYCAU3 were detected in horses for the first time. At locus ECAYA16, two alleles were found previously in Chinese indigenous breeds, namely Allele A (156 bp) and Allele B (152 bp) (Ling *et al.* 2010), and we found two new alleles of ECAYA16 in the present study. EcaYNO4 was polymorphic in western horse breeds, and a rare variant of the locus was carried by a Shetland pony (Kreutzmann *et al.* 2014). However, no polymorphism was detected at locus EcaYNO4 in this study. It was obvious that the horses in SWG had higher Y-STR PIC than horses from other regions. Although the population size of JJ horses was small, it had the highest diversity at locus ECAYA16. The networks showed that both Chinese indigenous and foreign horses share the most widespread young haplotypes, while there were relatively distant haplotypes harbored by horses from GZ, DBP, and JJ, distributed in Southwest China. The results were different from a previous report, in which Mongolia male horses from North China was found to have the deepest split in extant horses analyzed with Y-chromosomal SNPs (Han *et al.* 2019). The phylogenetic dendrogram indicated that the Chinese indigenous horses have a deep split with the introduced horses and clustered into two groups. However, the bootstrap values between some northern indigenous breeds are relatively low, which may be attributed to the small sample sizes of the breeds and low levels of divergence among them revealed with the Y-STRs. What is worthy of notice is that the blood relationship among our samples could not be completely excluded though we tried to avoid it during sampling, as pedigrees of the majority of Chinese indigenous horses have not been recorded. Blood relationship within a sampled population may cause the underestimation of the diversity. Again, if sample sizes of the studied breeds are similar, the comparison will be easier.

Table 3 Genetic diversity and PIC of the studied equine breeds[a]

Group	Breeds	Y-SNP			Y-STR PIC				
		π	h	n1	ECAYA16	ECAYP12	ECAYP13	ECAYCAU3	n2
MG	IMG	0.0003 ± 0.00002	0.286 ± 0.196	2	0	0	0	0	1
	ELC	0.00005 ± 0.00001	0.533 ± 0.095	3	0	0	0	0	1
KZKG	YQ	0	0	1	0	0	0.24	0	2
	KZK	0.00005 ± 0.00001	0.485 ± 0.106	2	0	0.41	0.21	0	4
		$(0.00004 \pm 1.4e\text{-}10)$	(0.447 ± 0.014)	(1.979 ± 0.021)					
HQG	CK	0.00003 ± 0.00002	0.417 ± 0.191	3	0	0.29	0	0	2
TIG	NQU	0.00003 ± 0.00002	0.286 ± 0.196	2	0.18	0.18	0.18	0	2
SWG	BS	0.00019 ± 0.00005	0.806 ± 0.089	4	0	0	0.16	0.27	3
	DBP	0.00016 ± 0.00002	0.844 ± 0.039	12	0.28	0.51	0.2	0.21	11
		$(0.00016 \pm 1.7e\text{-}09)$	(0.832 ± 0.008)	(5.515 ± 1.255)	(0.24 ± 0.019)	(0.44 ± 0.011)	(0.17 ± 0.012)	(0.16 ± 0.014)	(4.82 ± 1.170)
	LJ	0.00010 ± 0.00002	0.745 ± 0.049	7	0	0.2	0	0.05	4
		$(0.00009 \pm 1.1e\text{-}09)$	(0.725 ± 0.010)	(4.031 ± 0.788)	(0)	(0.16 ± 0.019)	(0)	(0)	(1.94 ± 0.597)
	TC	0.00008 ± 0.00001	0.710 ± 0.060	7	0.16	0.32	0	0	5
		$(0.00010 \pm 8.3e\text{-}10)$	(0.736 ± 0.010)	(4.293 ± 0.870)	(0.13 ± 0.016)	(0.28 ± 0.020)	(0)	(0)	(3.00 ± 0.711)
	GZ	0.00019 ± 0.00003	0.872 ± 0.054	6	0.3	0.62	0.35	0.14	7
		$(0.00018 \pm 1.3e\text{-}09)$	(0.806 ± 0.004)	(4.689 ± 0.673)	(0.27 ± 0.008)	(0.54 ± 0.009)	(0.31 ± 0.006)	(0.31 ± 0.005)	(5.22 ± 0.822)
	JJ	0.00010 ± 0.00006	0.511 ± 0.164	3	0.34	0	0.18	0	4
	NQ	0.00010 ± 0.00006	0.400 ± 0.237	2	0	0	0	0	1
AR	AR	0.00003 ± 0.00001	0.343 ± 0.128	3	0	0	0	0	1
		$(0.00004 \pm 4.2e\text{-}10)$	(0.328 ± 0.032)	(2.28 ± 0.424)					
TB	TB	0.00001 ± 0.00001	0.154 ± 0.126	2	0	0	0	0	1
		$(0.00001 \pm 6.1e\text{-}11)$	(0.177 ± 0.024)	(1.639 ± 0.231)					
AT	AT	0.0000 ± 0.0000	0	2	0	0	0	0	2
		$(0.00002 \pm 4.3e\text{-}11)$	(0.421 ± 0.017)	(1.97 ± 0.029)					
WB	WB	0.00003 ± 0.00001	0.343 ± 0.128	2	0	0	0	0	4

[a] π indicates nucleotide diversity; h, haplotype diversity; n1, number of haplotypes of Y-SNPs; PIC, polymorphic information content of Y-STRs; n2, number of haplotypes of Y-STRs. The values in brackets were obtained from resampling of size 10 with 1,000 replications to adjust for differences in sample size.
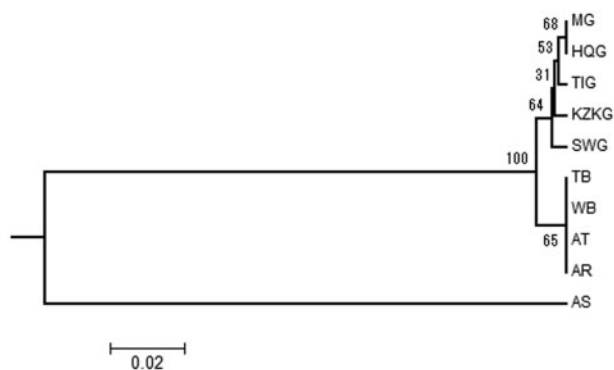
Figure 1 Dendrogram showing evolutionary relationships among the ten studied populations revealed with Y-STRs. The numbers at the nodes are values for 10,000 bootstrap replicates.

But the sample sizes of our studied population vary due to the difficulties to access some local breeds, especially male horses, and we had to normalize the sample sizes by resampling with a computer program. The introduced horses studied in the present study are from the popular breeds distributed across continents. Considering the relatively small sample size of introduced horses included in this study, further study with a large sample size, especially for the male horses, is still needed to verify our results. Despite these limitations, our results are still reliable, considering a relatively large total number of the indigenous samples, collected across the mainland of China.

Mitochondrial and Y-chromosomal polymorphism in Chinese indigenous horses

The results derived from our previous data showed that the northern indigenous horses (MG, KZKG, and HQG) had higher diversity than the southern horses (TIG and SWG) in mtDNA nucleotide and haplotype ($P < 0.05$). In contrast, the northern horses harbored lower haplotype diversity of Y chromosome than the southern horses ($P < 0.05$). The reasons underlying the phenomenon may exist in both horse immigration and geographic aspects. In China, the indigenous horses have been used primarily as labor animals for centuries and did not undergo any significant selection. However, there were still some recorded gene flows from the west in North China, such as the introgression of West Asian horses into northern China in the Han Dynasty, and some northern horses were crossed with the heavy horse breeds from the former Soviet Union in the 1950s, and the indigenous horses have been hybridized with Thoroughbreds introduced from western countries in the last two decades (Cui 1990; Chang 2011). As there are open grasslands in North China, these regions have been ideal places for horse raising from ancient times, attracting gene flows from adjacent places and forming a pool of maternal lines. On the other hand, as northern horses had been selected for military usage, stallions underwent stronger selection pressure than female horses (Lindgren *et al.* 2004). In contrast to North China, Southwestern China where most southern horses are distributed, is mountainous, which hindered gene flow, and the gene introgression and exchange of maternal lineages in the southern populations were much less than their northern counterparts. The local horses in the mountainous areas have been mainly bred as labor animals, and much less selection was imposed on the stallions compared with their northern counterparts, which led to more paternal lines preserved in South China.

Eneolithic Botai culture from Central Asia provides the earliest archaeological evidence of horse domestication, but Botai-like
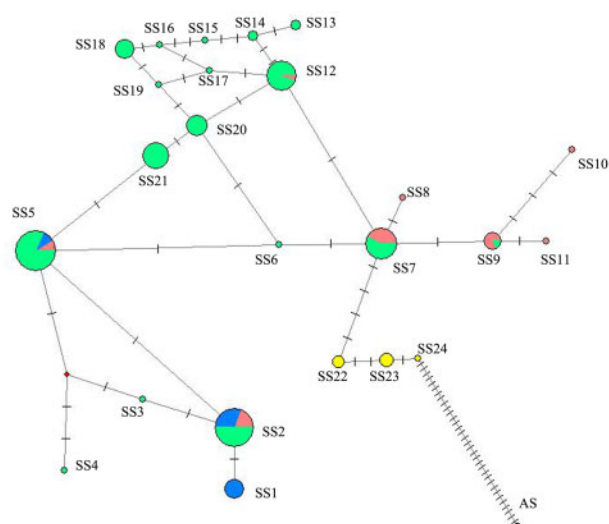

Figure 2 MJN network constructed using Y-STR and Y-SNP haplotypes. Circle size was proportional to the haplotype frequency. The number of mutations was shown in branches. Blue, introduced breeds; green, southern horses; red, northern horses; yellow, horses with *AMELY* mutations, and black indicates donkey.

horses were not the direct ancestors of modern horses, so the origins of the modern domestic horses have not been determined (Gaunitz et al. 2018; Fages et al., 2019). South American descendants of Iberian horses and Spanish horse breeds retained very ancient maternal lineages (Mirol *et al.* 2002), and later the Iberia Peninsula was regarded as a center of horse domestication (Lopes *et al.* 2005). The clear phylogeographical structure and the affinity of the southern horses to donkey may suggest that it could not be completely excluded that the events of horse domestication occurred in Southwest China. However, the presumption of the local domestication of horses certainly requires further evidence from both genetic and archaeological studies.

Application of the results from the maternal and paternal analyses on conservation of genetic diversity

In previous studies, Lei *et al.* (2009) and Ling (2010) revealed a significant phylogeographical association between horses of North China and haplogroup B of mtDNA, and the link between the southern horses and haplogroup G. In this study, distributions of haplogroups B and G also showed notable phylogeographical structures, but their frequencies ($B = 0.021$ and $G = 0.025$) decreased severely compared with those reported in previous studies about 10 years ago ($B = 0.043$ and $G = 0.043$) (Ling 2010). At present, the horses of haplogroup B have become rare in NQU, BS, DBP, LJ, TC, and JJ, and the same trend was also observed for the individuals belonging to haplogroup G in IMG, YQ, CK, NQU, TC, JJ, and NQ, which needs special attention regarding genetic conservation. In addition, several studies suggested that haplogroup F of mtDNA may originate from eastern Eurasia and spread to western Eurasia (McGahern *et al.* 2006; Cai 2007; Lei *et al.* 2009). In this study, the frequency of haplogroup F (0.192) in the indigenous breeds was lower than that reported 10 years ago (0.229) (Ling 2010), which indicated that the overall level of F was decreasing. The results suggest that more efforts should be made to protect the decreasing maternal lines in the indigenous breeds.

Most of the paternal lines are preserved in the southern indigenous horses. Some of the patrilines are rare, such as the males

with the AMILY-C allele and the 10 unique patrilines detected in the indigenous horses. As the total population size of the indigenous horses has decreased rapidly, detailed plan should be made to conserve the diversity of the horse patrilines. An ideal way to protect the paternal lines would be to conserve the individuals carrying the rare haplotypes by identifying with Y-chromosomal markers.

Breeds can be classified into three statuses (very safe, not at risk, and endangered) (Scherf 2000; Rischkowsky and Pilling 2007). JJ, NQ, and ELC, having a small population size less than 600 individuals, are in endangered status. Many endangered breeds demonstrate low levels of genetic diversity and have a small population size, but JJ showed relatively high genetic diversity, which may be attributed to the high level of the maternal lineage diversity (Ma *et al.* 2019). Urgent measures of conservation should be applied to JJ, NQ, and ELC due to their small population size of breeding stocks. The priority should be given to JJ, which is the only indigenous horse breed in the southeastern coastal region of China (Chang 2011), and also possesses some special paternal lines. Although other Chinese indigenous horse breeds are not at risk, more attention should be paid to LJ, TC, CK, and DBP than the rest of the breeds, which have relatively low-genetic diversity or limited population size.

The development of modern breeding has impacted genetic diversity more dramatically in the past 200 years, and led to severe decline of autosomal π diversity (Fages *et al.* 2019). The two aspects of breed extinction, loss of genetic diversity and loss of animals, are deeply interconnected. Future research should focus not only on endangered horse populations but also on disappearing genetic variants. In the DNA level, it is also needed to integrate the autosomal data into the current maternal and paternal genetic data pools in order to obtain more detailed evaluations on the genetic diversity of local breeds and make more comprehensive and efficient plan to preserve the indigenous genetic resources.

Conclusions

In conclusion, we revealed that Chinese indigenous horses possess rich polymorphisms of Y-chromosome, which suggested that there is higher paternal diversity preserved in the Chinese indigenous horse populations than expected. Prior conservation should be applied to the rare patrilineages and the decreasing maternal lines as well as the endangered indigenous breeds.

Acknowledgments

Funding

Literature cited

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

Brandariz-Fontes C, Leonard JA, Vega-Pla JL, Backstrom N, Lindgren G, *et al.* 2013. Y-chromosome analysis in Retuertas horses. PLoS One. 8:e64985.

Byun SO, Fang Q, Zhou H, Hickford JGH. 2009. An effective method for silver-staining DNA in large numbers of polyacrylamide gels. Anal Biochem. 385:174–175.

Cai DW. 2007. Ancient DNA and the Origins of Domestic Animals-the Research on the Origins of Chinese Domestic Sheep and Horses. P 63. Jilin University.

Chang H. 2011. Animal Genetic Resources in China-Horses, Donkeys, and Camels. Beijing: China Agricultural Press.

Cieslak M, Pruvost M, Benecke N, Hofreiter M, Morales A, *et al.* 2010. Origin and history of mitochondrial DNA lineages in domestic horses. PLoS One. 5:e15311.

Cui Y. 1990. Horse Science. 2nd ed. Beijing: China Agriculture Press.

Excoffier L, Lischer HE. 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.

Fages A, Hanghoj K, Khan N, Gaunitz C, Seguin-Orlando A, *et al.* 2019. Tracking five millennia of horse management with extensive ancient genome time series. Cell. 177:1419–1435.

FAO 2017. Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/en/#data.

Felkel S, Vogl C, Rigler D, Dobretsberger V, Chowdhary BP, *et al.* 2019. The horse Y chromosome as an informative marker for tracing sire lines. Sci Rep. 15:6095.

Felkel S, Vogl C, Rigler D, Jagannathan V, Leeb T, *et al.* 2018. Asian horses deepen the MSY phylogeny. Anim Genet. 49:90–93.

Gaunitz C, Fages A, Hanghoj K, Albrechtsen A, Khan N, *et al.* 2018. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science. 360:111–114.

Hamann H, Distl O. 2008. Genetic variability in Hanoverian warmblood horses using pedigree analysis. J Anim Sci. 86:1503–1513.

Han H, Wallner B, Rigler D, Machugh D E, Manglai D, *et al.* 2019. Chinese Mongolian horses may retain early domestic male genetic lineages yet to be discovered. Anim Genet. 50:399–402. 10.1111/age.12780

Han HY, Zhang Q, Gao KX, Yue XP, Zhang T, *et al.* 2015. Y-single nucleotide polymorphisms diversity in Chinese indigenous horse. Asian-Australas J Anim Sci. 28:1066–1074.

Hendricks BL. 2007. International Encyclopedia of Horse Breeds. University of Oklahoma Press.

Jansen T, Forster P, Levine MA, Oelke H, Hurles M, *et al.* 2002. Mitochondrial DNA and the origins of the domestic horse. Proc Natl Acad Sci USA. 99:10905–10910.

Jiang Q, Wei Y, Huang Y, Jiang H, Guo Y, *et al.* 2011. The complete mitochondrial genome and phylogenetic analysis of the Debao pony (*Equus caballus*). Mol Biol Rep. 38:593–599.

Kayser M, Roewer L, Hedman M, Henke L, Henke J, *et al.* 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet. 66:1580–1588.

Kreutzmann N, Brem G, Wallner B. 2014. The domestic horse harbours Y-chromosomal microsatellite polymorphism only on two widely distributed male lineages. Anim Genet. 45:460.

Lei CZ, Su R, Bower MA, Edwards CJ, Wang XB, *et al.* 2009. Multiple maternal origins of native modern and ancient horse populations in China. Anim Genet. 40:933–944.

Lenstra JA, Groeneveld LF, Eding H, Kantanen J, Williams JL, *et al.* 2012. Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. Anim Genet. 43: 483–502.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25: 1451–1452.

Lindgren G, Backström N, Swinburne J, Hellborg L, Einarsson A, *et al.* 2004. Limited number of patrilines in horse domestication. Nat Genet. 36:335–336.

Ling YH. 2010. Diversity and phylogenetic relationship of major local horse population in China (Doctoral thesis in Chinese). P83.Chinese Academy of Agricultural Sciences.

Ling YH, Ma YH, Guan WJ, Cheng YJ, Wang YP, *et al.* 2010. Identification of Y chromosome genetic variations in Chinese indigenous horse breeds. J Hered. 101:639–643.

Ling YH, Ma YH, Guan WJ, Cheng YJ, Wang YP, *et al.* 2011. Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers. Anim Genet. 42:56–65.

Lippold S, Knapp M, Kuznetsova T, Leonard JA, Benecke N, *et al.* 2011. Discovery of lost diversity of paternal horse lineages using ancient DNA. Nat Commun. 2:450.

Liu S, Yang Y, Pan Q, Sun Y, Ma H, *et al.* 2020. Ancient patrilineal lines and relatively high ECAY diversity preserved in indigenous horses revealed with novel Y-chromosome markers. Front Genet. 11:467.

Lopes MS, Mendonca D, Cymbron T, Valera M, da Costa-Ferreira J, *et al.* 2005. The Lusitano horse maternal lineage based on mitochondrial D-loop sequence variation. Anim Genet. 36:196–202.

Ma H, Wang S, Zeng G, Guo J, Guo M, *et al.* 2019. The origin of a coastal indigenous horse breed in china revealed by genome-wide SNP data. Genes (Basel). 10:241.

McGahern A, Bower MA, Edwards CJ, Brophy PO, Sulimova G, *et al.* 2006. Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. Anim Genet. 37: 494–497.

Mirol PM, Peral Garcia P, Vega-Pla JL, 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.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular-data.II. Gene-frequency data. J Mol Evol. 19: 153–170.

National Bureau of Statistics 2017. China Statistical Yearbook. http://www.stats.gov.cn.

Rafeie F, Amirinia C, Javaremi AN, Mirhoseini SZ, Amirmozafari N. 2011. A study of patrilineal genetic diversity in Iranian indigenous horse breeds. Afr J Biotechnol. 10:17347–17352.

Rischkowsky B, Pilling D. 2007. The State of the World's Animal Genetic Resources for Food and Agriculture. Rome: Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations.

Scherf BD. 2000. World Watch List for Domestic Animal Diversity. 3rd ed. Rome: Food and Agriculture Organization of the United Nations.

Vila C, Leonard JA, Gotherstrom A, Marklund S, Sandberg K, *et al.* 2001. Widespread origins of domestic horse lineages. Science. 291:474–477.

Wallner B, Brem G, Müller M, Achmann R. 2003. Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. Anim Genet. 34:453–456.

Wallner B, Palmieri N, Vogl C, Rigler D, Bozlak E, *et al.* 2017. Y Chromosome uncovers the recent oriental origin of modern stallions. Curr Biol. 27:2029–2035.

Wallner B, Piumi F, Brem G, Muller M, Achmann R. 2004. Isolation of Y chromosome-specific microsatellites in the horse and cross-species amplification in the genus Equus. J Hered. 95: 158–164.

Wallner B, Vogl C, Shukla P, Burgstaller JP, Druml T, *et al.* 2013. Identification of genetic variation on the horse Y chromosome and the tracing of male founder lineages in modern breeds. PLoS One. 8:e60015.

Wutke S, Sandoval-Castellanos E, Benecke N, Dohle HJ, Friederich S, *et al.* 2018. Decline of genetic diversity in ancient domestic stallions in. Sci Adv. 4:eaap9691.

Xu LX, Yang SL, Lin RY, Yang HB, Li AP, *et al.* 2012. Genetic diversity and population structure of Chinese pony breeds using microsatellite markers. Genet Mol Res. 11:2629–2640.

Yang Y, Zhu Q, Liu S, Zhao C, Wu C. 2017. The origin of Chinese domestic horses revealed with novel mtDNA variants. Anim Sci J. 88:19–26.

Yeh FC, Yang RC, Boyle TB, Ye Z, Mao JX. 1997. POPGENE, the user-friendly shareware for population genetic analysis. Mol Biol Biotechnol. 434:724–731.

Yue XP, Qin F, Campana MG, Liu DH, Mao CC, *et al.* 2012. Characterization of cytochrome b diversity in Chinese domestic horses. Anim Genet. 43:624–626.

Communicating editor: P. Morrell