Journal of Archaeological Science 36 (2009) 835–842

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Ancient DNA provides new insights into the origin of the Chinese domestic horse

Dawei Cai [a], Zhuowei Tang [a], Lu Han [b], Camilla F. Speller [c], Dongya Y. Yang [c], Xiaolin Ma [d], Jian'en Cao [e], Hong Zhu [a], Hui Zhou [a,b,*]

[a] Ancient DNA Laboratory, Research Center for Chinese Frontier Archaeology of Jilin University, Changchun 130012, China
[b] College of Life Science, Jilin University, Jiefang Road 115, Changchun 130023, China
[c] Ancient DNA Laboratory, Department of Archaeology, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A 1S6
[d] Henan Provincial Institute of Cultural Relics and Archaeology Zhengzhou 450000, China
[e] Inner Mongolian Institute of Cultural Relics and Archaeology Hohhot 010010, China

ARTICLE INFO

Article history:
Received 21 August 2008
Accepted 11 November 2008

Keywords:
Ancient DNA
Domestic horse
Mitochondrial DNA
Haplogroup
Przewalski horse

ABSTRACT

Domestic horses played a pivotal role in ancient China, but their exact origin remains controversial. To investigate the origin of Chinese domestic horses, we analyzed mitochondrial DNA (mtDNA) from 35 horse remains, aged between 4000 and 2000 years, excavated from nine archaeological sites in northern China. The Chinese ancient horses exhibited high matrilineal diversity, falling into all the seven haplogroups (A–G) observed in modern horses. These results suggest that several maternal lines were introduced into the gene pool of Chinese horses in the past. Haplogroups A and F were more prevalent in ancient horses than the other haplogroups. Interestingly, only haplogroups A and F were present in the samples older than 4000 years, while the more recent horses (between 2000 and 3000 years BP) fell into all seven haplogroups. Comparison with DNA data of present-day horses suggests that haplogroup F is like to be an ancient haplogroup of East Asian origin. These analyses also suggest that the origin of Chinese domestic horses is complex, and external mtDNA input occurred after initial domestication. Our results indicate that the Chinese ancient horses are more related to the modern Mongolian horses. Lastly, our results cannot support the previous hypothesis that early Chinese domestic horses were derived from the Przewalski horse.

1. Introduction

Humans have had a profound influence on the horse since its domestication in the late Neolithic. Used for transport, labour, food and recreation, horses became important in many facets of our society (Mills and McDonnell, 2005). Archaeological evidence suggests that the horse was probably first domesticated about 6000 years ago on the grassland steppes of Eurasia from Ukraine to Turkestan, and the earliest domesticated horses subsequently spread from this area (Anthony, 1996; Bennett and Hoffmann, 1999; Clutton-Brock, 1999). However, mitochondrial DNA (mtDNA) analysis of modern and ancient horses has revealed extensive variation within and among breeds, with little congruence of haplogroup distribution to breeds or geographic areas, suggesting that separate and geographically diverse populations participated in domestication process, and that domestic horses have multiple origins (Jansen et al., 2002; Lister et al., 1998; Vila et al., 2001).

Domestic horses played a pivotal role in ancient China, being used mainly in agriculture and transport, especially in military purposes (Chen, 1994). However, the geographic origin of Chinese domestic horses remains controversial. Zooarchaeological data show that abundant domestic horse remains appeared suddenly in China at the sites of the Shang Dynasty (ca. 3000 BP). Prior to the Shang Dynasty, there are few records of domestic horses. Excavation of thousands of Neolithic and early Bronze sites in China yielded only a few sporadic fragments of horse tooth and bone at a limited number of sites, and it is difficult to determine whether the remains are those of wild or domestic horses (Chen, 1999). The absence of evidence for the early stages of domestication, coupled with the sudden appearance of horses around 3000 BP, have led scholars to suggest two hypotheses. First, that domestic horses were imported into China from the Eurasian steppe via that Gansu and Qinhai provinces (Yuan and An, 1997), or, second, that horses underwent autochthonous domestication in China during the Later Neolithic (ca. 4000 BP) or perhaps even earlier, and Przewalski horses were proposed as the ancestors of Chinese domestic horses (Wang and Song, 2001; Zhang, 2004). In order to discriminate between these two hypotheses, we used ancient DNA techniques to

* Corresponding author. College of Life Science, Jilin University, Jiefang Road 115, Changchun 130023, China. Tel./fax: +86 431 8849 8031.
E-mail address: zhouhui@mail.jlu.edu.cn (H. Zhou).

0305-4403/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jas.2008.11.006

investigate the genetic diversity and phylogenetic relationships of Chinese ancient horses, as this approach has proved useful for the study of the origins and spread of other livestock (Cai et al., 2007a,b; Leonard et al., 2002; McGahern et al., 2006; Troy et al., 2001).

We analyzed mitochondrial DNA (mtDNA) sequences from 26 ancient horses excavated from different archaeological sites in Northern China, along with nine previously reported sequences from Mongolia (Cai et al., 2007a). The sequences were compared with previously published ancient, modern and wild horses stored in the GenBank.

2. Materials and methods

2.1. Archaeological information

The ancient horse samples came from seven archaeological sites in Northern China (Fig. 1): three from the Inner Mongolia Autonomous Region (two sites – Bancheng and Xiaoshuanggucheng – in Liangcheng County, and Xindianzi in Helingeer County); two from Henan Province (Maoyuan in Xinzheng City, Yin Ruins in Anyang City); Qianzhangda site in Tengzhou City, Shandong Province; Yujiazhuang site in Pengpu Village, Guyuan County, Ningxia Hui Autonomous Region. The ages of these sites vary from ca. 3000 BP (Shang Dynasty) to ca. 2500 BP (from the end of Spring–Autumn to the start of the Warring States period) (Table 1).

2.2. Archaeological horse remains

Thirty-seven horse teeth and bones were retrieved from seven archaeological sites [Inner Mongolia: Bancheng ($n=7$), Xiaoshuanggucheng ($n=4$), Xindianzi ($n=13$); Henan Province: Maoyuan ($n=8$), Yin Ruins ($n=2$); Qianzhangda ($n=1$) and Yujiazhuang ($n=2$)] (Table 1). After morphological studies, the samples were stored in sealed plastic bags placed in boxes in a store room. Prior to their arrival at the ancient DNA laboratory, the remains had been handled by less than five archaeozoologists. The samples were of different ages, and exhibited different macroscopic preservation. Most of the tooth samples from Inner Mongolia were well preserved, with smooth and compact external surfaces. In contrast, most of the horse bones from Henan Province were light-weight and sandy coloured, with damaged external surfaces.

Fig. 1. Geographical location of archaeological sites. ① Bancheng and Xiaoshuanggucheng; ② Xindianzi; ③ Dashanqian; ④ Jinggouzi; ⑤ Yin Ruins; ⑥ Maoyuan; ⑦ Qianzhangda; and ⑧ Yujiazhuang.

Table 1
Archaeological horse samples studied, with associated codes, skeletal element used, dates and PCR results.

Lab code	Archaeological code	Skeletal element	Date	No. of extractions	Result
Bancheng site					
LB01	03LBM4-1	Tooth	ca. 2500 BP	2	Reproducible
LB02	03LBM20-1	Tooth	ca. 2500 BP	2	Reproducible
LB03	03LBM26-1	Tooth	ca. 2500 BP	2	Reproducible
LB04	03LBM28-1	Tooth	ca. 2500 BP	2	Reproducible
LB05	03LBM31-1	Tooth	ca. 2500 BP	2	Reproducible
LB06	03LBM40-3	Tooth	ca. 2500 BP	2	Reproducible
LB07	03LBM67-2	Tooth	ca. 2500 BP	2	Reproducible
Xiaoshuanggucheng site					
LS01	03LSM4-2	Tooth	ca. 2500 BP	2	Reproducible
LS02	03LSM7-4	Tooth	ca. 2500 BP	2	Reproducible
LS03	03LSM9-10	Tooth	ca. 2500 BP	2	Reproducible
LS04	03LSM11-1	Tooth	ca. 2500 BP	2	Reproducible
Xindianzi site					
HX01	99HXM2-2	Tooth	ca. 2500 BP	2	Reproducible
HX02	99HXM2-3	Tooth	ca. 2500 BP	2	Reproducible
HX03	99HXM2-7	Tooth	ca. 2500 BP	2	Reproducible
HX04	99HXM5-2	Tooth	ca. 2500 BP	2	Reproducible
HX05	99HXM7-1	Tooth	ca. 2500 BP	3	Reproducible
HX06	99HXM20-1	Tooth	ca. 2500 BP	2	Reproducible
HX07	99HXM34-1	Tooth	ca. 2500 BP	2	Reproducible
HX08	99HXM37-7	Tooth	ca. 2500 BP	2	Reproducible
HX09	99HXM38-2	Tooth	ca. 2500 BP	2	Reproducible
HX10	99HXM43-2	Tooth	ca. 2500 BP	3	Reproducible
HX11	99HXM43-9	Tooth	ca. 2500 BP	3	PCR fail
HX12	99HXM43-10	Tooth	ca. 2500 BP	3	Reproducible
HX13	99HXM43-16	Tooth	ca. 2500 BP	2	Reproducible
Qianzhangda site					
QZ01	#1	Tooth	ca. 3000 BP	4	PCR fail
Yujiazhuang site					
YJZ01	87YJZ11	Tooth	ca. 2500 BP	2	Reproducible
YJZ02	87YJZ20	Tooth	ca. 2500 BP	2	Reproducible
Maoyuan site					
MY01	03ZHMY21	3rd phalange	ca. 2500 BP	3	PCR fail
MY02	03ZHMY22	3rd phalange	ca. 2500 BP	3	PCR fail
MY03	03ZHMY23	3rd phalange	ca. 2500 BP	3	PCR fail
MY04	03ZHMY24	3rd phalange	ca. 2500 BP	3	PCR fail
MY05	03ZHMY25	3rd phalange	ca. 2500 BP	4	Reproducible
MY06	03ZHMY26	3rd phalange	ca. 2500 BP	2	PCR fail
MY07	03ZHMY27	3rd phalange	ca. 2500 BP	2	PCR fail
MY08	03ZHMY28	3rd phalange	ca. 2500 BP	3	PCR fail
Yin Ruins					
YX01	YXH433	2nd phalange	ca. 3000 BP	2	PCR fail
YX02	YXHF26	2nd phalange	ca. 3000 BP	3	Contradiction

To increase the sample size, we included nine previously reported sequences, (DQ900922–DQ900930) from nine horse remains excavated from Dashaqian ($n=5$) and Jinggouzi ($n=4$) sites, Chifeng region, Inner Mongolia, China, dating from the Lower Xiajiadian Culture (ca. 4000 BP) to the late Warring States period (ca. 2000 BP) (Cai et al., 2007a).

2.3. DNA extraction, amplification and sequencing at Jilin University

Before DNA extraction, bone and tooth samples were prepared using the procedure described by Cai et al. (2007a). DNA was extracted following a silica-spin column method (Yang et al., 1998). Several DNA extractions were performed for each specimen as shown in Table 1.

A 300 bp fragment of the mtDNA D-loop region (nucleotides 15473–15772 based on reference sequences X79547) was amplified by using two sets of overlapping primers described by Cai et al. (2007a): forward primer L15473 5′–CTTCCCCTAAACGACAACAA–3′ and reverse primer H15692 5′–TTTGACTTGGATGGGGTATG–3′; and forward primer L15571 5′–AATGGCCTATGTACGTCGTG–3′ and reverse primer H15772 5′–GGGAGGGTTGCTGATTTC–3′. PCR amplifications were performed in an Mastercycler® personal Thermal Cycler (Eppendorf, Hamburg, Germany) in a 50 μL reaction volume containing 2.5 mM Mg^{2+}, 1 × Buffer, 0.2 mM dNTPs, 1.6 mg/ml BSA, 0.5 μM of each primer, 1 U Taq polymerase (Promega, USA) and 2 μL DNA template. The PCR condition were as follows: predenaturation at 95 °C for 5 min, 8 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min; followed by 28 cycles of 92 °C for 1 min, 50–55 °C for 1 min, 72 °C for 1 min; and a final elongation step of 72 °C for 10 min.

The PCR products were isolated from 2% agarose gels and purified by using QIAEX® GEL Extraction Kit (QIAGen, Germany). Both DNA strands of PCR products were directly sequenced with original primers on an ABI 310 automated DNA sequencer (Applied Biosystems, USA). The obtained electropherograms were assembled to examine any base pair ambiguities using Chromas 2.22 (www.technelysium.com.au).

2.4. DNA extraction, amplification and sequencing at Simon Fraser University

Three bone samples (LB04, LS02 and HX07) were extracted using a similar DNA extraction method (Yang et al., 1998) in the dedicated ancient DNA laboratory at Simon Fraser University. PCR amplifications were conducted in a Mastercycler Personal in a 30 μL reaction volume containing 50 mM KCl, 10 mM Tris–HCl, 2.5 mM $MgCl_2$, 0.2 mM dNTP, 1.0 mg/mL BSA, 0.3 μM each primer, 3.0 μL DNA sample and 1.5–3.0 U AmpliTaq Gold™ (Applied Biosystems).

The above sets of primers (the Jilin lab) were used for PCR amplifications. PCR was run for 60 cycles at 94 °C for 30 s (denaturing), 52–55 °C for 30 s (annealing), and 72 °C extension for 40 s, with an initial 12 min denaturing period at 95 °C. Five microliters of PCR product were visualized via electrophoresis on a 2% agarose gel using SYBR Green™ staining. PCR products were purified using QIAGEN's MinElute™ purification kits and were subjected to direct sequencing through Macrogen (www.macrogen.com). PCR products were sequenced using both reverse and forward primers, and electropherograms for the same DNA sample were assembled and cross-examined using Chromas Pro (www.technelysium.com.au).

2.5. Comparison sequences

Ancient horse sequences were truncated to 247 bp (15494–15740) for phylogenetic analysis, and compared with 1053 modern horse mtDNA sequences from GenBank, including 17 populations from five broad geographical regions (Table S1): East Asia: Tuva ($n = 11$), Cheju ($n = 34$), Mongolian ($n = 33$), Guan Mountain ($n = 10$), Tibetan ($n = 16$); North Asia: Vyatskaya ($n = 18$), yakut ($n = 20$); Central Asia: Akhal Teke ($n = 24$), Orlov ($n = 18$); Middle East and Africa: Turkey ($n = 27$), Arabia ($n = 87$), Africa ($n = 20$); Europe: Iberia ($n = 251$), British Isles ($n = 223$), Mainland Europe ($n = 218$), Northern Europe [Mesenskaya ($n = 18$), Scandinavia ($n = 25$)]. Finally, we added the wild Przewalski horse (AF014409, AF072994–AF072995, AJ413830–AJ413832 and AF326635) and ancient horse sequences from Ireland (DQ327848, DQ327850–DQ327851), Kazakhstan (AJ876883–AJ876890), Russia (AJ876891–AJ876892), Korea (AF049720) and northern Europe (AF326676–AF326679), and late Pleistocene sequences from Alaska (AF326668–AF326675).

2.6. Phylogenetic analysis

Multiple alignments were performed using the Clustal X 1.83 program (available at ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) (Thompson et al., 1997). The median-joining (MJ) network (Bandelt et al., 1999) was drawn using the program NETWORK4.1.1.2 (available at www.fluxus-engineering.com/sharenet.htm). Haplotype diversity, nucleotide diversity, mismatch analysis, Fu's F_S values and Linearised F_{ST} distances were computed using ARLEQUIN 2.0 software (available at http://cmpg.unibe.ch/software/arlequin/) (Schneider and Excoffier, 2000). The statistical significance of the values was estimated by permutation analysis using 1000 replications.

3. Results

3.1. Sequence variation and haplogroups assignment

Ancient DNA sequences were recovered from 26 of the 37 samples. Ten samples (HX11 from Xindianzi site, MY01-04 and MY06-08 from Maoyuan site, QZ01 from Qianzhangda site and YX01 from Yin Ruins) did not yield PCR products. In addition, one sample YX02 from Yin Ruins was excluded because sequences obtained from several extractions were different. The samples from Inner Mongolia showed the higher rate of success than those from other regions, especially Henan, which can be attributed to the cold and dry environment of Inner Mongolia which is more suited to DNA survival. Ancient horse sequences data in this paper have been submitted to GenBank with accession numbers: EU931584–EU931609.

The 35 ancient horses (including the nine sequences reported in Cai et al., 2007a and 26 sequences reported here) yielded 35 variable positions when compared with the reference sequence (GenBank X79547). All the substitutions were transitions, and defined 25 different haplotypes (H1–H25, Table 2). Eighteen haplotypes were singletons, while seven haplotypes were detected more than once, of which three haplotypes (H6, H8 and H13) were shared by samples from the same archaeological site, and four haplotypes (H3, H5, H9 and H19) between different sites. The 25 haplotypes were assigned to all of the seven haplogroups A–G in modern horses, identified by Jansen et al. (2002) and in the skeleton network described by McGahern et al. (2006) (Table 1).

3.2. Haplogroup frequency

Haplogroups A and F occurred most frequently in our study (42.8% and 31.4%, respectively). Haplogroups B, C, D, E and G were found at lower frequencies (2.9%, 5.7%, 2.9%, 8.6% and 5.7%). We compared our results with the frequency of the seven haplogroups in 17 modern horse populations and Przewalski horses (Table 3 and Fig. 2). Haplogroup A is the most common haplogroup in most modern horse populations, with the highest frequency (66.6%) in Orlov to the lowest (22%) in Mesenskaya. Interestingly, all analyzed Przewalski horses belong to haplogroup A. Haplogroup F has the highest frequency in Mongolian (36.4%) and Cheju horses (32.4%), and the lowest frequency in Iberian (2.8%) and British Isles horses (3.2%).

3.3. Genetic diversity

Within-species genetic diversity is thought to reflect population size, history, ecology, and the ability to adapt, and can be assessed by calculating the haplotype diversity and nucleotide diversity within a population. In this study, the haplotype diversity and nucleotide diversity of Chinese ancient horses were 0.978 ± 0.012 and 0.025 ± 0.013, respectively. Interestingly, the genetic diversity was very similar in the modern horses (Table 3). The Przewalskis

Table 2
Variable nucleotides and lineage assignment of 35 ancient horses.

Haplotypes	Variable nucleotide positions																																	Lineages	Samples
	15494	15495	15496	15521	15526	15534	15538	15540	15542	15585	15595	15596	15597	15598	15601	16002	16003	16004	16015	16116	16356	16496	16505	16596	16606	16667	17003	17019	17118	17202	17233	17266	17373	17407		
X79547	T	T	A	G	T	C	A	A	C	G	A	A	A	T	T	C	T	G	A	A	C	A	A	T	A	G	A	T	C	C	G	C	G	T	A		
H1		C		A								G				T															A			C		E	DQ900927
H2		C							T							T					T				A		C			A					A	DQ900928
H3		C														T		A					C				C			A		A		G	F	DQ900929, HX10
H4		C									G		G			T		A									C			A				G	F	DQ900930
H5		C											G			T		A								G	C			A					F	DQ900924, HX07, LB07
H6		C							T	A						T		A					G			A				A					A	DQ900922, DQ900923, DQ900925
H7		C																					G			A		C			A				A	DQ900926
H8		C											G			T		A			T						G	C		T	A				F	HX01, HX03
H9		C							T	A			G			T					T		G			A		C			A				A	HX02, HX06, LB01
H10		C																														T			A	HX04
H11		C														T		A							G			C		T	A		A		G	F	HX05
H12		C														T			G	G			C					C			A					G	HX08
H13		C		A												T															A			C		E	HX09, HX13
H14		C											G			T							G								A					A	HX12
H15		C			C			G		A						T						G							T		A				A	LB02
H16		C								A					C	T															A				C	LB03
H17	C	C	G			T										T	C					G									A				D	LB04
H18		C													C	T															A				C	LB05
H19		C														T															A				A	LB06, MY05
H20		C							T				G			T					T		G			A		C			A				A	LS01
H21		C								A						T															A				A	LS02
H22		C												C		T			G				C					C			A				G	LS03
H23		C														T												C			A			G	F	LS04
H24		C					G			A						T													T		A				B	YJZ01
H25		C														T		A										C			A			G	F	YJZ02

had the lowest values (0.476 ± 0.171 and 0.0019 ± 0.0017, respectively) reflecting the severe genetic bottleneck suffered by these animals.

3.4. Phylogenetic network construction

In order to investigate the phylogenetic relationship of the haplotypes, we constructed a median-joining (MJ) network with 1121 sequences of ancient, modern and wild horses (Fig. 3). Seven major haplogroups and 17 clusters (A1–A7, B1–B2, C1–C2, D1–D3 and F1–F3) were evident in the star-like network. The 25 haplotypes of our study fell into all of the seven haplogroups, and in several of the clusters. Some haplotypes, for example H5 and H25, shared the founder haplotype of a cluster with a variety of breeds from different regions. The network showed that breeds of different geographic regions overlapped, suggesting extensive gene flow had occurred among different breeds. Interestingly, some haplogroups corresponded to particular breeds and regions. For example, haplogroup F clustered with breeds from East Asia and Middle East, while D corresponded with many Iberian horses. Several ancient

Table 3
Diversity indices and frequency of mtDNA lineages in 19 horse populations.

Breed/population	Code	Number of individuals	Number of haplotypes	Haplotype diversity (SE)	Nucleotide diversity (SE)	Lineage frequency (%)						
						A	B	C	D	E	F	G
Ancient horses	Anc	35	25	0.9782 ± 0.0120	0.0246 ± 0.0133	42.8	2.9	5.7	2.9	8.6	31.4	5.7
Mongolian	Mon	33	22	0.9564 ± 0.0205	0.0293 ± 0.0157	24.2	6.1	12.1	21.2	–	36.4	–
Guan Mountain	Gua	10	8	0.9556 ± 0.0594	0.0258 ± 0.0151	40.0	–	30.0	–	–	20.0	10.0
Tibetan	Tib	16	14	0.9833 ± 0.0278	0.0273 ± 0.0152	50.0	–	–	18.8	18.7	12.5	–
Korea Cheju	Kor	34	20	0.9554 ± 0.0182	0.0250 ± 0.0136	38.2	17.6	5.9	–	5.9	32.4	–
Tuva	Tuv	11	11	1.000 ± 0.0388	0.0325 ± 0.0185	45.4	9.1	–	18.2	–	18.2	9.1
Yakut	Yak	20	12	0.9421 ± 0.0295	0.0244 ± 0.0135	55.0	–	–	25.0	–	20.0	–
Vyatskaya	Vya	18	10	0.9281 ± 0.0373	0.0211 ± 0.0120	61.1	–	–	33.3	–	5.6	–
Mesenskaya	Mes	18	11	0.9150 ± 0.0500	0.0225 ± 0.0127	22.2	27.8	22.2	5.6	11.1	11.1	–
Orlov	Orl	18	9	0.8693 ± 0.0610	0.0209 ± 0.0119	66.6	16.7	11.1	5.6	–	–	–
Arabia	Ara	87	43	0.9767 ± 0.0050	0.0259 ± 0.0137	41.4	16.1	4.6	17.2	–	19.5	1.2
Akha-teke	Akh	24	14	0.9312 ± 0.0326	0.0278 ± 0.0151	29.2	4.2	12.5	29.2	8.2	12.5	4.2
Turkey	Tur	27	18	0.9687 ± 0.0172	0.0232 ± 0.0127	29.6	7.4	18.5	29.6	–	14.8	–
Africa	Afr	20	17	0.9789 ± 0.0245	0.0300 ± 0.0164	50.0	5.0	–	40.0	–	5.0	–
Iberia	Ibe	251	70	0.9419 ± 0.0082	0.0274 ± 0.0143	38.6	6.4	8.0	41.8	0.8	2.8	1.6
British Isles	Bri	223	79	0.9721 ± 0.0035	0.0286 ± 0.0149	23.8	8.5	20.6	34.5	9.4	3.2	–
Scandinavia	Sca	25	13	0.9233 ± 0.0300	0.0241 ± 0.0133	24.0	4.0	36.0	12.0	4.0	20.0	–
Maniland Europe	Eur	218	87	0.9712 ± 0.0041	0.0297 ± 0.0154	32.5	7.8	10.6	31.2	0.5	7.8	9.6
Przewalskii	Prz	7	2	0.4762 ± 0.1713	0.0019 ± 0.0017	100	–	–	–	–	–	–

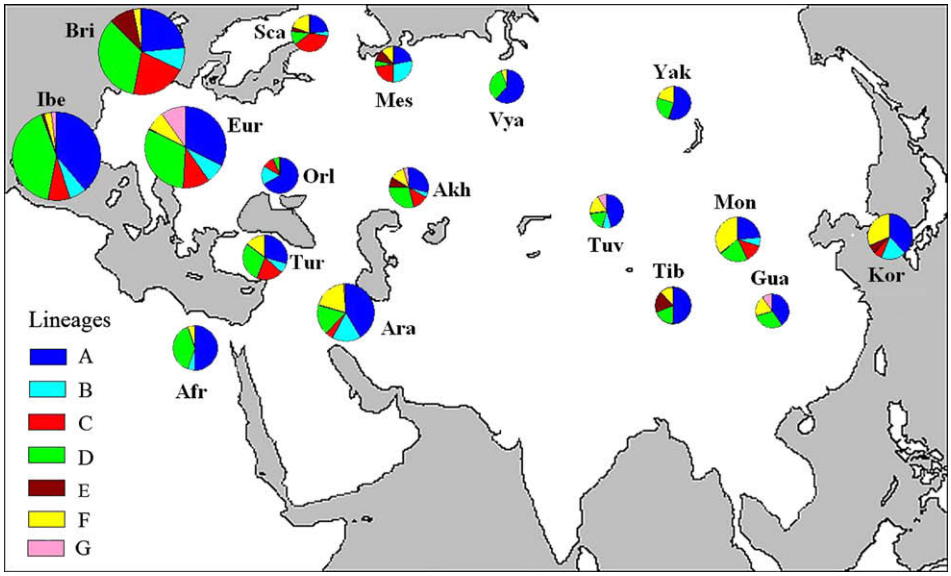

Fig. 2. Hose mtDNA lineages distributions in 17 horse populations.

horses from other archaeological sites also fell into the seven haplogroups, suggesting that the seven haplogroups have ancient origins.

3.5. Population expansion

The historical demography of populations has a profound effect on patterns of genetic variation (Chen et al., 2006). Mismatch distribution analysis and Fu's *Fs* statistic can be used to reveal the genetic signature of population expansion. In this study, the mismatch analysis for 35 ancient horses showed a bell-shaped distribution (figure not shown), and an additional test performed by Fu's *Fs* statistics gave a significant negative value (−25.26, $P = 0.0000$), suggesting that Chinese ancient horses had undergone population expansion events.

3.6. Genetic distance

Linearised F_{ST} values can be used as an estimate of genetic distances between populations over shallow time depths (Slatkin,

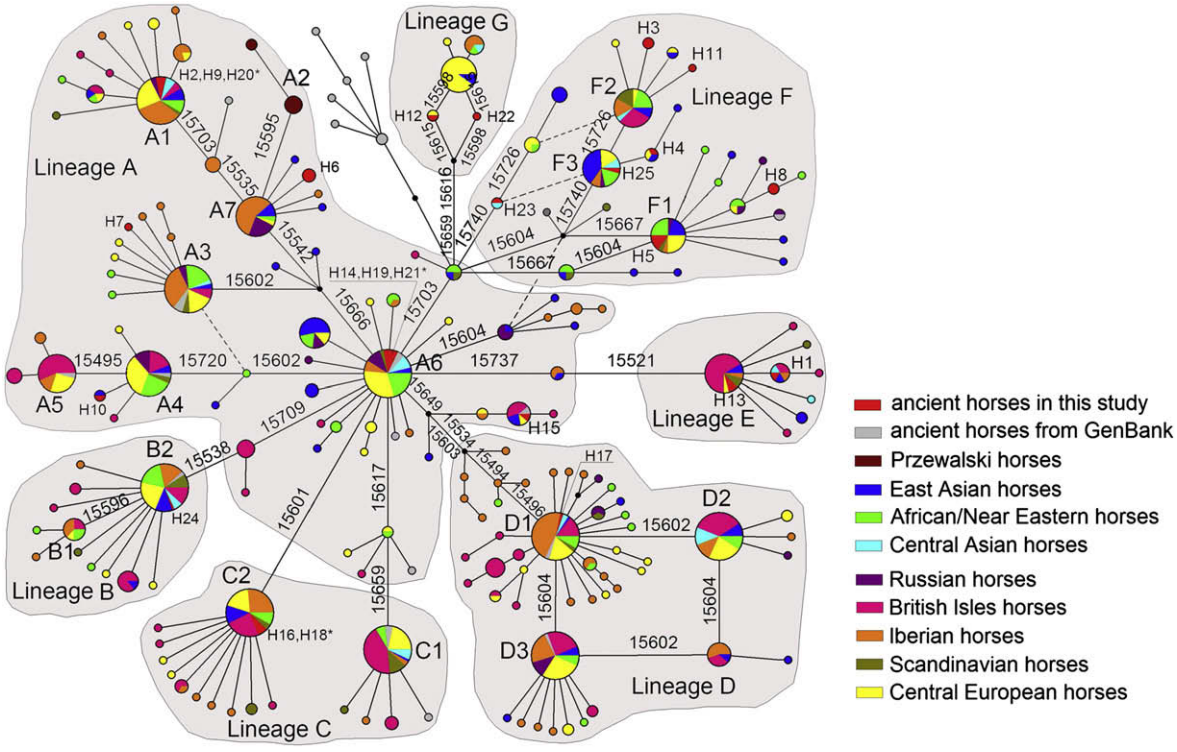

Fig. 3. Phylogenetic networks of 1121 domestic horses, wild horses and ancient horses. Circle areas are proportional to mtDNA haplotype frequencies. mtDNA lineage and clusters named according to Jansen et al. (2002) and McGahern et al. (2006). *Mutational hotspots (15585, 15597, 15650) were excluded.

Table 4
Linearised F_{ST} diatance (below diagonal) and the significant of F_{ST} *P* values between 19 populations.

	Anc	Mon	Gua	Tib	Kor	Tuv	Yak	Vya	Mes	Orl	Ara	Akh	Tur	Afr	Ibe	Bri	Sca	Eur	Prz
Anc		+	+	+	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+
Mon	0.0336		−	−	+	−	−	−	+	+	−	−	−	−	+	+	−	−	+
Gua	0.0756	0.0000		−	+	−	−	−	−	+	−	−	−	−	−	−	−	−	+
Tib	0.0486	0.0115	0.0000		−	−	−	+	−	+	−	−	−	−	+	+	−	+	+
Kor	0.0213	0.0409	0.0634	0.0342		−	+	+	−	+	+	+	+	+	+	+	+	+	+
Tuv	0.0000	0.0000	0.0000	0.0000	0.0125		−	−	−	−	−	−	−	−	−	−	−	−	+
Yak	0.0570	0.0083	0.0000	0.0000	0.0691	0.0000		−	−	+	−	−	−	−	+	+	+	−	+
Vya	0.1458	0.0387	0.0271	0.0806	0.1369	0.0000	0.0387		+	+	+	−	−	−	−	−	+	−	+
Mes	0.0541	0.0433	0.0876	0.0451	0.0000	0.0133	0.0982	0.1365		−	−	−	+	+	+	+	−	+	+
Orl	0.1259	0.0763	0.1396	0.1149	0.0908	0.0419	0.1246	0.0896	0.0614		+	+	+	+	+	+	+	+	+
Ara	0.0413	0.0032	0.0327	0.0217	0.0311	0.0000	0.0224	0.0382	0.0155	0.0402		−	+	+	+	+	+	+	+
Akh	0.0668	0.0002	0.0000	0.0055	0.0745	0.0000	0.0274	0.0327	0.0463	0.0890	0.0158		−	−	−	−	−	−	+
Tur	0.1194	0.0077	0.0000	0.0459	0.0964	0.0000	0.0332	0.0000	0.0858	0.0818	0.0294	0.0000		0.0107	+	−	−	−	+
Afr	0.1375	0.0334	0.0157	0.0375	0.1416	0.0000	0.0222	0.0000	0.1124	0.1257	0.0389	0.0086	0.0269	0.0000	−		−	−	+
Ibe	0.1449	0.0471	0.0389	0.0525	0.1545	0.0180	0.0397	0.0247	0.1171	0.1275	0.0498	0.0135	0.0000	0.0241	0.0425				
Bri	0.1552	0.0522	0.0000	0.0617	0.1308	0.0222	0.0509	0.0152	0.0907	0.0970	0.0636	0.0112	0.0211	0.0795	0.0886	0.0480			
Sca	0.0488	0.0191	0.0164	0.0318	0.0456	0.0000	0.0477	0.0725	0.0405	0.0821	0.0296	0.0093	0.0039	0.0086	0.0194	0.0212	0.0306		
Eur	0.0831	0.0182	0.0000	0.0346	0.0904	0.0000	0.0247	0.0169	0.0605	0.0838	0.0235	0.0000	0.0039	0.0086	0.0194	0.0212	0.0306	0.6875	0.5125
Prz	0.5822	0.5639	1.1510	0.6514	0.6546	0.5773	0.6955	0.8939	0.8645	0.8254	0.5051	0.6607	0.8235	0.5921	0.4940	0.6407	0.6875	0.5125	

1995). In this study, the genetic distances between 19 populations were assessed by linearised F_{ST} values (Table 4). The Chinese ancient horses showed the lowest values, with horses from East Asia (Tuva, Cheju and Mongolian horses), while the highest values were found in the Przewalski (0.5822, $P < 0.05$). Generally, the genetic distances between Chinese ancient horses and the 17 modern populations increased from East to West. These results suggest that the Chinese ancient horses have an affinity to East Asian horses.

4. Discussion

4.1. Authenticity of ancient DNA sequences

A major concern in studying DNA recovered from archaeological material is the authenticity of the results. One advantage of working with animal rather than human bone is that species-specific primers, which exclude the amplification of human DNA, can be used for PCR (Barnes and Young, 2000). Therefore, the most likely forms of contaminations are cross-contamination between ancient samples, or PCR product carry-over. In this study, we took the following precautions to ensure that our results were genuine and not the result of contamination: (1) all experiments were performed in a dedicated ancient DNA laboratory in which modern horse DNA had not been previously isolated or analyzed; (2) strict contamination controls were exercised throughout all steps according to commonly accepted recommendations (Cooper and Poinar, 2000), and no contamination was detected in the negative controls of DNA extraction or PCR amplification; (3) sequences were replicated for each sample from multiple independent extractions (Table 1). Furthermore, several samples were validated independently at a separate ancient DNA laboratory, located in the Department of Archaeology at Simon Fraser University, Canada. (4) To verify the accuracy of direct sequencing, PCR products of several samples (HX03, LB07, LS03 and YJZ01) were cloned using the PGemT Easy cloning kit (Promega) according to the handbook. Five clones were picked for sequencing. The consensus sequence from the clones was compared with that from direct sequencing, with both types of sequences having the same sequence, indicating that the sequences generated by direct sequencing were authentic.

4.2. The origins of Chinese horses

Modern horses exhibit high mtDNA diversity, indicating that the maternal lineages have a widespread origin. In the present study, we also observed high diversity in the Chinese ancient horses. We found 25 haplotypes belonging to all seven haplogroups A–G found in modern horses. Some of the ancient horses also shared the founder haplotype of a cluster with a variety of breeds of different geographical regions. This suggests that several different maternal lines were introduced into the gene pool of the Chinese horses in the past. To determine the origins of Chinese domestic horses, we need to learn the time and place of the origin of the gene pool. Haplogroups A and F are more abundant than the other haplogroups. Interestingly, haplogroups A and F were the only haplogroups detected in samples (DQ90022–DQ90024, Cai et al., 2007a) aged more than 4000 years, while the samples aged between 2000 and 3000 years had all haplogroups. These results indicate a multiple origin scenario, where horses belonging to haplogroups A and F were first domesticated in China more than 4000 years ago, after which time the horses of the other haplogroups were introduced, consistent with a population expansion.

To shed light on the origin of the diverse haplogroups, we studied the distribution of the seven haplogroups in 17 modern populations, and showed that only F and D have a significant geographical pattern (Table 3 and Fig. 2). Haplogroup F is prevalent

in East Eurasian populations and its frequencies decline from east to west, consistent with the results of previous studies (McGahern et al., 2006). Based on our findings in Chinese horses older than 4000 years, we propose that haplogroup F is an ancient haplogroup of East Asian origin. Haplogroup D is abundant in Western Eurasian populations, especially Iberia (42%) and its frequencies decline from west to east. Considering the abundant archaeological evidence from the Iberia Peninsula and previous studies, we suggest that haplogroup D originated in the Iberian Peninsula. We did not find a clear geographic specificity for haplogroups A, B, C, E and G, although previous studies suggest that E and C tend to correspond to ponies from the western and northern fringes of Europe (McGahern et al., 2006; Jansen et al., 2002). Additional data are required to reveal the geographical origins of these haplogroups.

In view of the proposed origin of haplogroups F and D, and the presence of diverse maternal lineages, the population expansion of Chinese ancient horses and the spread of the knowledge of horse breeding about 1250 BC (Levine, 1999), we suggest that the origin of Chinese domestic horses is more complex than previously thought, and that both indigenous breeds and introduced maternal lineages were involved in the process of domestication. Thus, our data fail to support either of the two previous hypotheses on the domestication of horses in China.

Generally, the genetic diversity (haplotype diversity and nucleotide diversity) in Chinese ancient horses was similar to that observed modern horse populations (Table 3), suggesting that Chinese ancient horses have an affinity to modern horse populations. Based on the present-day distribution of haplogroup F and genetic distance analysis, the Chinese ancient horses appear to be most closely related to the Mongolian horses and Cheju horses (Tables 3 and 4). The Cheju horses are thought to be of Mongolian origin as historical records indicate that the Mongols first introduced 160 horses to Cheju Island in 1276 (Nam, 1969). In conclusion, Chinese ancient horses are closely related to the Mongolian horses.

4.3. Ancient horses and Przewalski horses

Przewalski horses were once widespread in the Eurasian steppe. As the Earth warmed following the last Ice Age, they lost ground and were confined to small areas in northern China and southern Mongolia. Many Przewalski fossils have been found in the late Pleistocene faunas and human settlements in northern China, from Xinjiang in the west, to the Taiwan Strait (Deng, 1999). Due to their widespread distribution, some scholars suggested that early Chinese domestic horses were derived from the Przewalski horse (Deng, 2000; Zhang, 2004).

Despite having a different chromosome number, Przewalski horses ($n = 66$) can interbreed with domestic horses ($n = 64$) to produce fertile offspring (Ryder et al., 1978; Trommerhausen-Smith et al., 1979). However, the phylogenetic analysis of Przewalski and domestic horses based on mtDNA and Y chromosome markers revealed two different genetic patterns. Using mtDNA sequences, Przewalski and domestic horses do not form monophyletic clades in a phylogenetic tree (Ishida et al., 1995; Oakenfull et al., 2000; Oakenfull and Ryder, 1998), while based on the Y chromosome, there were two fixed nucleotide differences between Przewalski horse and domestic horse (Wallner et al., 2003). These studies suggested that female mediated gene flow occurred between domestic horse and the Przewalski horse.

In our study, we detected no maternal connection between the domestic horse and the Przewalski. Phylogenetic analysis revealed that the Przewalski fell into cluster A2, distinct from Chinese ancient horses and modern horses. Genetic distance analysis indicated that the Przewalski's were far from the ancient horses, suggesting that the Przewalski horse and the domestic horse should be considered sister taxa with a common ancestor, and that the Przewalski horse is not the wild ancestor to the Chinese domestic horse. However, this conclusion should be viewed with caution, as the Przewalski horse underwent a severe genetic bottleneck: all present individuals descend from only 13 survivors, with only four maternal lineages (Volf and Kus, 1991). A large-scale analysis of the ancient Przewalski horse DNA would permit the detection of potential extinct maternal lineages.

5. Conclusion

In conclusion, our results reveal the genetic diversity and phylogenetic structure of ancient Chinese domestic horses. This research provides valuable new insights into the origin of Chinese domestic horse: (1) haplogroup F is most likely to be an ancient haplogroup of East Asian origin; (2) the origin of Chinese domestic horses is complex, and external mtDNA input occurred during domestication; and (3) Chinese ancient horses were more related to the modern Mongolian horses. As a result, our results do not support the previous hypothesis that early Chinese domestic horses were derived from the Przewalski horse.

Acknowledgements

The work was supported by the National Science Fund for Fostering Talents in Basic Research (grant no. J0530184). We are grateful to Dr Erika Hagelberg for correcting grammar and helpful comments on the manuscript. We also thank Henan Provincial and Inner Mongolian Institute of Cultural Relics and Archaeology for providing samples for analysis.

Supplementary information

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jas.2008.11.006.

References

Anthony, D.W., 1996. In: Olsen, S.L. (Ed.), Horses Through Time. Roberts Rinehart for Carnegie Museum of Natural History, Boulder, CO, pp. 57–82.
Bandelt, H.J., Forster, P., Rohl, A., 1999. Median joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16, 37–48.
Barnes, I., Young, J.P.W., 2000. DNA-based identification of goose species from two archaeological sites in Lincolnshire. Journal of Archaeological Science 27, 91–100.
Bennett, D., Hoffmann, R.S., 1999. *Equus caballus*. Mammalian Species 628, 1–14.
Cai, D.W., Han, L., Xie, C.Z., Li, S.N., Zhou, H., Zhu, H., 2007a. Mitochondrial DNA analysis of Bronze Age horses recovered from Chifeng region, Inner Mongolia, China. Progress in Natural Science 17, 544–550.
Cai, D.W., Han, L., Zhang, X.L., Zhou, H., Zhu, H., 2007b. DNA analysis of archaeological sheep remains from China. Journal of Archaeological Science 34, 1347–1355.
Chen, W.H., 1994. Agricultural Archaeology Catalogue of China. Jiangxi Scientific & Technical Publisher, Jiangxi (in Chinese).
Chen, X.C., 1999. The origins of Chinese horses. Chinese Cultural Relic Newspaper (in Chinese).
Chen, S.Y., Duan, Z.Y., Sha, T., Xiangyu, J., Wu, S.F., Zhang, Y.P., 2006. Origin, genetic diversity, and population structure of Chinese domestic sheep. Gene 376, 216–223.
Clutton-Brock, J., 1999. A Natural History of Domesticated Mammals, second ed. Cambridge University Press, Cambridge.
Cooper, A., Poinar, H.N., 2000. Ancient DNA: do it right or not at all. Science 289, 1139.
Deng, T., 1999. Historical distrbution of *Equus przewalskii* and its control by climate fluctuation. Acta Theriologica Sinica 19, 10–16.
Deng, T., 2000. Phylogenetic relationship of the Chinese miniature pony to *Equus przewalskii* (Perissodacatyla, Equidae). Acta Veterinaria et Zootechnica Sinica 31, 28–33 (in Chinese).
Ishida, N., Oyunsuren, T., Mashima, S., Mukoyama, H., Saitou, N., 1995. Mitochondrial DNA sequences of various species of the genus Equus with special reference to the phylogenetic relationship between Przewalski's wild horse and domestic horse. Journal of Molecular Evolution 41, 180–188.
Jansen, T., Forster, P., Levine, M.A., Oelke, H., Hurles, M., Renfrew, C., Weber, J., Olek, K., 2002. Mitochondrial DNA and the origins of the domestic horse. Proceedings of the National Academy of Sciences of the United States of America 99, 10905–10910.

Leonard, J.A., Wayne, R.K., Wheeler, J., Valadez, R., Guillén, S., Vila', C., 2002. Ancient DNA evidence for Old World origin of New World dogs. Science 298, 1613–1616.
Levine, M.A., 1999. Botai and the origins of horse domestication. Journal of Anthropological Archaeology 18, 29–78.
Lister, A.M., Kadwell, M., Kaagan, L.M., Jordan, W.C., Richard, M.B., Stanley, H.E., 1998. Ancient and modern DNA in a study of horse domestication. Ancient Biomolecules 2, 267–280.
McGahern, A.M., Edwards, C.J., Bower, M.A., Heffernan, A., Park, S.D.E., Brophy, P.O., Bradley, D.G., MacHugh, D.E., Hill, E.W., 2006. Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. Animal Genetics 37, 498–502.
Mills, D.S., McDonnell, S.M., 2005. The Domestic Horse: The Origins, Development and Management of Its Behaviour. Cambridge University Press.
Nam, D.Y., 1969. Horse production in Cheju during Lee dynasty. In: Studies on Korean History, vol. 4. Korea History Research Society, College of Letters and Sciences, Seoul National University, Seoul, South Korea, pp. 77–131.
Oakenfull, E.A., Ryder, O.A., 1998. Mitochondrial control region and 12S rRNA variation in Przewalski's horse (*Equus przewalskii*). Animal Genetics 29, 456–459.
Oakenfull, E.A., Lim, H., Ryder, O., 2000. A survey of equid mitochondrial DNA: implications for the evolution, genetic diversity and conservation of *Equus*. Conservation Genetics 1, 341–355.
Ryder, O., Epel, N.C., Benirschke, K., 1978. Chromosome banding studies of Equidae. Cytogenetics and Cell Genetics 20, 323–350.
Schneider, S., Excoffier, L., 2000. Arlequin: A Software for Population Genetics Data Analysis. Ver 2.000. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139, 457–462.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25, 4876–4882.
Trommerhausen-Smith, A., Ryder, O., Suzuki, Y., 1979. Bloodtyping studies of twelve Przewalski's horses. International Zoo Yearbook 19, 224–227.
Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C., Bradley, D., 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature 410, 1088–1091.
Vila, C., Leonard, J.A., Gotherstrom, A., Marklund, S., Sandberg, K., Liden, K., Wayne, R.K., Ellegren, H., 2001. Widespread origins of domestic horse lineages. Science 291, 474–477.
Volf, J., Kus, E., 1991. General Studbook of the Przewalski Horse. Zoological Garden Prague, Prague, 308 pp.
Wallner, B., Brem, G., Mullerller, M., Achmann, R., 2003. Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. Animal Genetics 34, 453–456.
Wang, Z.J., Song, P., 2001. Exploring the origination of domesticated horses in Northern China. Archaeology and Cultural Relics 2, 26–30 (in Chinese).
Yang, D.Y., Eng, B., Waye, J., Dudar, J.C., Saunders, S.R., 1998. Technical Note: improved DNA extraction from Ancient Bones using silica-based spin columns. American Journal of Physical Anthropology 105, 539–543.
Yuan, J., An, J.Y., 1997. Two important questions of Chinese Zooarchaeological Research. Chinese Cultural Relic Newspaper (in Chinese).
Zhang, C.S., 2004. Wild horses, domestic horses, and center for Eastern Asia horse raising. Agricultural Archaeology 1, 252–254 (in Chinese).