Complex Evolutionary Patterns Revealed by Mitochondrial Genomes of the Domestic Horse

T. Ning[1,2,3], J. Li[4,5], K. Lin[6], H. Xiao[7], S. Wylie[8], S. Hua[9], H. Li[*,6] and Y.-P. Zhang[*,1,3]

[1]*Laboratory for Conservation and Utilization of Bio-resource and Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Yunnan, China*

[2]*Laboratory for Animal Genetic Diversity and Evolution of Higher Education in Yunnan Province, Yunnan University, Yunnan, China*

[3]*State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Yunnan, China*

[4]*Agriculture College, Kunming University, Yunnan, China*

[5]*Engineering Research Center for Urban Modern Agriculture of Higher Education in Yunnan Province, Kunming University, Yunnan, China*

[6]*CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China*

[7]*School of Life Sciences, Yunnan University, Yunnan, China*

[8]*Western Australian State Agricultural Biotechnology Centre, Murdoch University, Perth, Australia*

[9]*Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Yunnan, China*

Abstract: The domestic horse is the most widely used and important stock and recreational animal, valued for its strength and endurance. The energy required by the domestic horse is mainly supplied by mitochondria *via* oxidative phosphorylation. Thus, selection may have played an essential role in the evolution of the horse mitochondria. Besides, demographic events also affect the DNA polymorphic pattern on mitochondria. To understand the evolutionary patterns of the mitochondria of the domestic horse, we used a deep sequencing approach to obtain the complete sequences of 15 mitochondrial genomes, and four mitochondrial gene sequences, *ND6*, *ATP8*, *ATP6* and *CYTB*, collected from 509, 363, 363 and 409 domestic horses, respectively. Evidence of strong substitution rate heterogeneity was found at nonsynonymous sites across the genomes. Signatures of recent positive selection on mtDNA of domestic horse were detected. Specifically, five amino acids in the four mitochondrial genes were identified as the targets of positive selection. Coalescent-based simulations imply that recent population expansion is the most probable explanation for the matrilineal population history for domestic horse. Our findings reveal a complex pattern of non-neutral evolution of the mitochondrial genome in the domestic horses.

Keywords: Horse, mitochondria, population history, recombination, selection.

INTRODUCTION

The domestication of animals played a major role in man's shift from a hunter-gatherer lifestyle to an agricultural one [1]. The horse (*Equus ferus caballus* L.) is believed to have first been domesticated in the Eurasian Steppes [2], although we believe independent horse domestication also occurred in East Asia (unpublished). The horse was particularly important because it served as transport, as a stock animal to pull plows *etc*, and as food - milk and meat. While the horse transformed people's lives, artificial selection by humans deeply influenced horse evolution. The many breeds in existence today are evidence that selective pressure has impacted on horse morphology and energy metabolism. The domestication process of the horse is to serve human beings with its strength and endurance. Notably, characteristics like strength and endurance are influenced by energy metabolism, and it is highly challenging to meet the energy requirement for the potential working load. Thus, we suspect that genes, organelles, organs or tissues involved in metabolism pathways may have experienced adaptive evolution.

*Address correspondence to these authors at the (H. Li) CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China;
Tel: 86-21-54920460; Fax: 86-21-54920451;
E-mail: lihaipeng@picb.ac.cn and (Y.-P. Zhang), State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China;
Tel: 86-871-65190761; Fax: 86-871-65195430;
E-mail: zhangyp@mail.kiz.ac.cn

1566-5240/14 $58.00+.00

The mitochondrion is a key organelle related to metabolism. It produces 95% energy for eukaryotic cells and encodes 13 proteins which encode four protein complexes in mammals: NADH dehydrogenase (*ND1*, *2*, *3*, *4*, *4L*, *5* and *6*), cytochrome bc1 (*CYTB*), cytochrome c oxidase (*COX1*, *2* and *3*) and ATP synthase (*ATP6* and *ATP8*). These complexes are involved in the oxidative phosphorylation (OXPHOS) process, which is essential for energy metabolism [3]. In addition to the 13 mitochondrial-encoded proteins, more than 76 nuclear genes encoding proteins work in concert to generate a functional OXPHOS system [4]. Both nuclear and mitochondrial proteins may modify metabolic fitness by functional interactions [5-7]. The co-evolution of the mitochondrial and nuclear-encoded OXPHOS genes has been demonstrated to occur in orangutan, marine copepod, pulmonate snail, human *etc* [8-12]. However, the metabolic properties of mitochondria make them highly mutagenic environments for the mtDNAs that lie within them [13]. Previous studies showed that mitochondrial mutations improved aerobic capacity to adapt to environments in anthropoid primate, tribe Caprini and carnivorous plant [4, 14, 15], and resulted in a large number of diseases of human [16-18], including oxidative phosphorylation diseases [19]. Thus, we hypothesize that selection on some mutations in the horse mtDNA may induce changes in metabolic performance. For a long time the mitochondrial genome was believed to evolve nearly neutrally. However, recent studies have challenged the view [20-24]. Evidences for non-neutral evolution [25, 26], paternal inheritance [27-30] and recombination [31-34] were also found in animal mtDNAs, although they are still considered exceptionally rare event. These reports suggest that mitochondrial evolution is more complex in some species than once thought.

The pattern of mtDNA evolution is driven by both selection and/or demographic events over the period of domestication. Domestic animals are characterized by a short demographic history, plentiful phenotypes, strong artificial selection, extinction of most wild ancestors and polygamous reproduction [35], all of which makes it difficult to research whether the conventional pattern of mtDNA evolution is appropriate for them. Little attention has been devoted to researching the role of artificial selection on the evolution of mtDNA in domestic animals. Previous researches reveal that the domestic dog (*Canis lupus familiaris*) and the domestic yak (*Bos grunniens* L.) have accumulated nonsynonymous changes in mitochondrial genes at a faster rate than their ancestors. These suggest that domestication have caused the relaxation of selective constraints on their mitochondrial genomes [36, 37]. Based on analysis of three mitochondrial genomes of the Tibetan plateau horse, the mitochondrial *ND6* gene was reported to accumulate larger numbers of nonsynonymous substitutions [38]. Subsequently, our team reveal that high-altitude environment has directed adaptive evolution of the mitochondrial *ND6* gene in the plateau horse through population genetic analysis [39].

To investigate whether selection played a role in the mtDNA evolution of the domestic horse, we conducted neutrality tests on two different data sets. One is based on DNA polymorphisms within complete mitochondrial genomes (n = 20, 15 new sequences and five existing ones), and the other is based on polymorphisms from four mitochondrial gene sequences, *ND6*, *ATP8*, *ATP6* and *CYTB*, collected from 509, 363, 363 and 409 individuals, respectively. To determine the probable demographic history of domestic horses, we calculated different summary statistics based on the two data sets and fitted those summary statistics with the models by large scale of coalescent-based simulations. Thus, this study reveals a detailed evolutionary pattern and the effect of positive selection on the mitochondrial genome of the domestic horse.

MATERIALS AND METHODS

Sampling and DNA Extraction

We focused on studying the evolutionary driving forces of mitochondrial genomes in East Asian horses. A total of 509 horses representing 24 pure Chinese breeds were collected from small remote villages belonging to ten geographic regions in China. Details of breeds, geographic regions, and sample sizes are given in Table **S1**. Genomic DNA was extracted from unrelated domestic horses using a standard phenol-chloroform method. Information about the horses from which the samples were collected was gathered from the owners of the animals and local farmers. We only took samples from well-defined populations [40].

DNA Amplification and Sequencing

Mitochondrial genomes were amplified and sequenced as described previously [38]. Four mitochondrial protein-coding genes (*ND6*, *ATP8*, *ATP6* and *CYTB*) were amplified using primers N6F (5'-CCAAAATCTA TCTCCCAGTT-3'), N6R (5'-GAGCCGATTTCATCAT-3'), A86F (5'-GACTTTACTACGGTCAATGCT -3'), A86R (5'-GTTTCCCTATTAGGCTATGGT-3'), CBF (5'-ATGATATGAAAAAC CATCGTTG-3') and CBR (5'-TCTCCTTCTCTGGTTTACAAGAC-3'). Amplification by PCR was done in a 50 μl reaction mixture containing approximately 50 ng genomic DNA, 5μl 10× PCR reaction buffer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 10 pmol each primer, and 1U *Taq* DNA polymerase (TaKaRa Biosystems). The reaction mixture was initially denatured at 95˚C for 3 min followed by 35 cycles of denaturation at 95˚C for 30s, annealing at either 58˚C (for mitochondrial genome, *ATP8* and *ATP6*) or 50˚C (for *ND6* and *CYTB*) for 1 min, and strand extension of 72˚C for 1.5 min. The PCR products were purified using a Watson™ PCR Purification Kit (Watson BioTechnologies) and both strands of each product were sequenced using ABI PRISM BigDye™ Terminator v3.1 technology (Applied Biosystems) on an ABI PRISM 3730 DNA Analyzer.

The raw sequences were edited within DNAStar package (DNASTAR, Madison, WI) and aligned with the Clustal W program [41]. The nucleotide position

(np) numbering follows that of a previously published horse complete mtDNA sequence (GenBank accession X79547) [42]. Possible nuclear mitochondrial pseudogenes (Numts) were eliminated based on a careful examination of sequence characteristics, including indels, in-frame stop codons, and nucleotide composition. Novel sequences were deposited in GenBank under accession numbers FJ718996-8, GU734774-85, FJ643628-75, FJ765096-208 and EU433643-86.

Data Analysis

Descriptive Analysis for Molecular Population Genetics

Since our focus was on the evolutionary forces shaping mitochondrial genomes in East Asian horses, we used four mitochondrial genome sequences of domestic horse from East Asia available in GenBank (EF597512-14, EU939445) and a standard reference sequence (X79547). Our preliminary phylogenetic analysis (with different tree-building methods) showed that a single East Asian horse sequence (AY584828) exhibits an unusually long branch, and severely departures from the hypothesis of clocklike evolution (results not shown). It might indicate contamination of the sequences by nuclear mitochondrial pseudogenes. Therefore, it was excluded from our analysis. Totally, 15 new horse mitochondrial genomes were used for this study along with five other sequences available on GenBank. We used a donkey (*Equus africanus asinus* L.) mitochondrial genome (X97337) as the outgroup.

The general characteristics of horse mitochondrial genomes were revealed using the DAMBE program [43]. A sliding window analysis was done by DnaSP v5 [44]. The window size was 500 nucleotides (nt), and the step size 100 nt. The linear correlation analysis between the summary statistics of windows and the nucleotide positions of windows was carried out by SPSS (SPSS Inc., version 13.0, Chicago IL).

The synonymous substitution rate (Ks), nonsynonymous substitution rate (Ka) and the ratio of nonsynonymous substitution rate to synonymous substitution rate (Ka/Ks) were calculated under the MS model of maximum likelihood method by KaKs_Calculator [45]. DnaSP v5 [44] was implemented to calculate the nucleotide diversity, Tajima's *D* [46], Fu & Li's *D** and *F** [47], Fu's *Fs* [48], the number of synonymous (S) and nonsynonymous (N) nucleotide substitution. The significant p values of the Tajima's *D*, Fu's *Fs* and Fu & Li's *D** and *F** were estimated under 10^4 coalescent simulations without recombination or by Fisher exact test. To ensure independent and conservative observations, correction values are not used in the MK test.

Tests of Neutrality and Selection

McDonald and Kreitman (MK) test [49] was performed by using DnaSP v5 [44]. To elucidate whether substitutions might be affected by positive selection in a particular region of a gene, we adopted TMpred software [50] to divide a gene into surface and transmembrane domains, and then survey selective pressures on the two parts by neutrality tests.

To scrutinize selection on mitochondrial protein-coding genes in the horse, we adopted the tree-based relative ratio test by CRANN [51], and likelihood approaches by PAML4 [52] and ADAPTSITE1.6 [53]. The expectation of the tree-based relative ratio test is that the ratio of replacement (RV) to silent (SV) polymorphisms on each branch should be equal to the ratio of replacement (RI) to silent (SI) substitutions over the entire tree under neutrality. The G test or the Fisher exact test was used to check significant differences between SV/RV to SI/RI. Moreover, the site-specific likelihood test was run under the CODEML module in PAML package. Two pairs of PAML site-specific likelihood models were compared: M1 (nearly neutral) and M2 (positive selection), M7 (beta) and M8 (beta & ω). The significance of twice the difference in maximum likelihood values between nested models was checked with the chi-square distribution. To identify candidate positive selection sites, Bayes empirical Bayes (BEB) analysis was utilized [54]. In addition to the PAML, the ADAPSITE program was used to infer positively selected amino acid sites.

Coalescent Simulation

Coalescent simulations were conducted with no recombination by conditioning on the observed sample size and the Watterson's θ (θ_w). Four demographic models (bottleneck, expansion, island and splitting models, schematic diagrams were shown in Fig. **S1**) are performed using the program ms [55]. The bottleneck model is specified by the parameters *F* (the inbreeding coefficient) [56] and *t* (the time in years measured from the present) at which the bottleneck occurred. The generally accepted date for the earliest known domestic horse is 5,500 years ago [2]. If we take effective population size N_0 = 320,000 and 4 years per generation [57], *t* is equal to 0.001 (=6000/4*4*320,000) in units of $4N_0$ generations. Values of *F* considered were *F* = [-0.00001, -0.0001, …, -0.1]. The expansion model is determined by the parameters α (the growth rate/generation) and *t* (the time in years measured from the present) at which the population began increasing in size. Values considered for α and *t* were: α = [0.001, 0.01 ..., 1000] and *t* = [1,000, 2,000, ..., 5,000]. The population structure under an island model is specified by the population migration rate between demes, $M = 4N_0m$, where N_0 and m are the effective subpopulation size and fraction of migrants in each subpopulation per generation, respectively. Values of *M* considered were *M* = [1, 2, ..., 10]. Population splitting model is specified with no subsequent migration and determined by the parameter *t* (the time in years since populations diverged). Values of *t* considered were *t* = [1,000, 2,000, . . ., 5,000].

Identification of Recombinations and Multiple Hits

To avoid method dependence, we detected evidence of recombination by utilizing six widely used indirect tests, including RDP [58], GENECONV [59],

BootScan [60], SiScan [61], MaxChi and Chimaera [62] as implemented in RDP2 [63]. The RDP2 program was used to survey the occurrence of recombination as well as identify particular recombination events. The p value cutoff and the number of permutations were set to 0.05 and 1,000 for all tests, respectively. Statistics of linkage disequilibrium which are generally used to detect recombination were not implemented because they tend to confound either the absence (|*D*'|) or presence (r^2) of recombination with a lack of power [64].

To check multiple hits (*i.e.* two or more mutations occurring at the same nt site), all mutations of horse mtDNA were located on a genealogy with Fitch's parsimony algorithm [65].

RESULTS

Genetic Diversity and Substitution

Fifteen new complete mitochondrial genome sequences of domestic horses were determined, as well as PCR-generated sequences from 509 samples of the *ND6* gene, 363 of *ATP8*, 363 of *ATP6* and 409 of *CYTB*.

Three summary statistics of genetic diversity (nucleotide diversity, θ per site and the number of polymorphic sites) were calculated based on the 20 mitochondrial genome sequences of domestic horse (15 new sequences and five existing ones) (Fig. **1**). It is noteworthy that there was a slightly significant positive correlation between diversity and nucleotide position. Among 15 mitochondrial genes, the genetic diversity of *16S rRNA* gene is the lowest. Moreover the nucleotide diversity (π) of the four complexes involved in oxidative phosphorylation decreased in the order of mitochondrial ATPs > NDs > COXs > CYTB in domestic horse (Fig. **S2**). To place the context of genetic diversity domestic horse amongst other large draft domestic animals, we compared nucleotide diversity based on mitochondrial genomes of five species: horse, donkey, yak, cattle, and buffalo (Table **S2**). Nucleotide diversity in the horse is second only to that of the buffalo (Fig. **S3**).

The observed and expected distributions (expected distributions were generated from the total observed polymorphisms multiplied by the proportional size of the gene) of synonymous substitutions on mitochondrial genomes of domestic horses is not significantly different among 13 protein-coding mitochondrial genes ($p = 0.514$) (Fig. **2A**). However, the observed distribution of nonsynonymous substitutions is extremely significantly different from the expected distribution among the 13 genes ($p < 0.001$) (Fig. **2B**). Notably, this extreme significance level ($p < 0.001$) remains even if the effect of rate variation among nonsynonymous substitutions was adjusted by multiplying the proportion of observed synonymous substitutions in each mitochondrial gene.

The relative rates of nucleotide substitution for the different mitochondrial genes are reported in Table **1**. Ks varies slightly across the 13 protein-coding genes. The highest Ks is only about 2-fold that of the lowest Ks. However, Ka varies markedly. The *ND6* and *ND3* genes have the highest and the lowest values, respectively. Moreover, the Ka/Ks ratios differ among 13 protein-coding genes. The *ND4*, *ND6*, *ATP6* and *ATP8* genes indicate an elevated Ka. Interestingly, the horse *ND6* and *ATP8* genes are characterized by high substitution rate variation resulting from both a higher Ka and a lower Ks. Specifically, the *ND6* gene exhibited the highest Ka/Ks ratios. Since both nonsynonymous and synonymous substitution rates are increased in the *ND4* and *ATP6* genes, their Ka/Ks ratios are not obviously elevated. We also found that the Ka/Ks ratios in the horse population are higher than those between the horse and donkey for all mitochondrial genes except the *ND4L* and *ATP8* genes. Moreover, the *ND6* and *ATP8* genes show the higher Ka/Ks ratio, while the lower Ka/Ks ratio was always detected in *COX* genes.

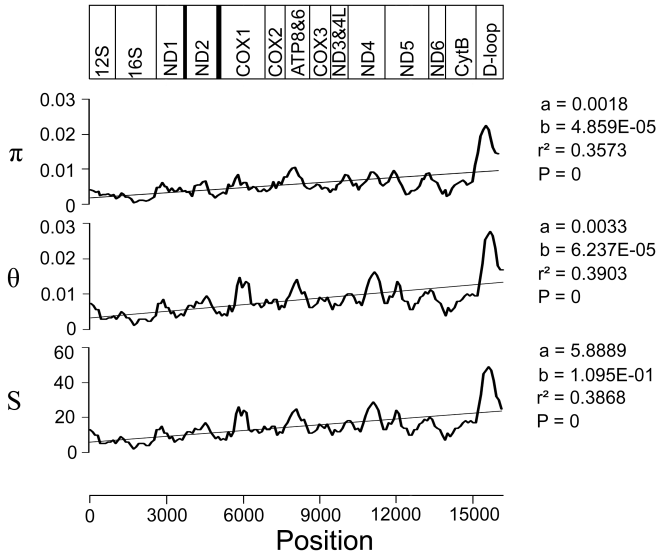

Fig. (1). DNA sequence polymorphism and allele frequency distribution by nucleotide position. The nucleotide diversity (π), θ per site and number of polymorphic sites (S) were calculated for a sliding window of 500 nucleotides (nt) with a step size of 100 nt, respectively. The regression line is represented by a straight line and its p value is assessed by *F* statistic. The parameters of regression formula (y = a + bx) and the p values are indicated on the right side of each sliding window.

Detection of Selection

Interestingly, eleven of thirteen mitochondrial genes exhibited neutral indices (NI) > 1 in the interspecies comparison between horse and donkey by the McDonald-Kreitman test (Table **2**). When we combine all the 13 genes together, NI = 2.908 which is very significantly larger than 1 ($p < 0.001$) (Table **2**). Thus, we can reject the hypothesis that the horse mtDNA protein-coding regions have evolved under the strictly neutral model. However, MK tests are pairwise, it is difficult to infer whether departures from neutrality were from one or both species [66].

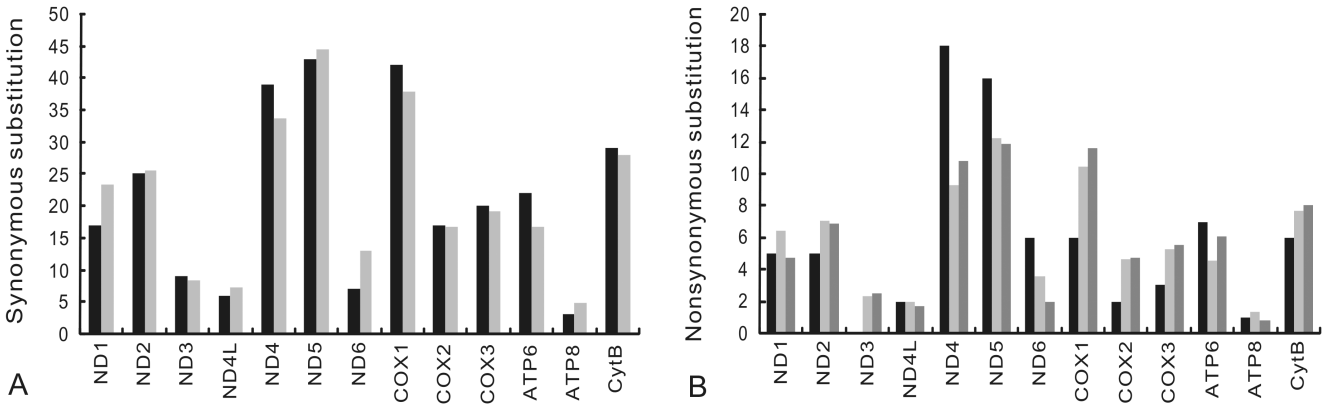

Fig. (2). Observed and expected synonymous and nonsynonymous substitutions. A: synonymous substitution; **B**: nonsynonymous substitution; the black bar indicates the observed value, the light grey bar is the expected value, and the dark grey bar is the adjusted value.

Table 1. Substitution rates of interspecies and intraspecies comparisons.

	Horse/Donkey				Horse			
	Ks	**Ka**	**Ka/Ks**	**Ks**	**Ka**	**Ka/Ks**	***α***	**I**
ND1	0.3089	0.0089	0.029	0.0150	0.0014	0.095	0.1013	0.602
ND2	0.3886	0.0199	0.051	0.0163	0.0012	0.074	0.0890	0.589
ND3	0.2974	0.0039	0.013	0.0182	0	0	0.2917	0.595
ND4L	0.3214	0.0183	0.057	0.0229	0.0009	0.041	0.2620	0.662
ND4	0.3856	0.0074	0.019	0.0219	0.0023	0.103	0.1128	0.538
ND5	0.4531	0.0130	0.029	0.0214	0.0014	0.063	0.1255	0.522
ND6	0.2216	0.0226	0.102	0.0148	0.0029	0.193	0.0851	0.662
COX1	0.3438	0.0011	0.003	0.0201	0.0006	0.028	0.0723	0.419
COX2	0.3787	0.0021	0.006	0.0210	0.0004	0.019	0.0700	0.485
COX3	0.3647	0.0021	0.006	0.0184	0.0008	0.043	0.0940	0.549
ATP6	0.4139	0.0138	0.033	0.0321	0.0024	0.074	0.1867	0.534
ATP8	0.2174	0.0483	0.222	0.0160	0.0026	0.160	0.0705	0.605
CYTB	0.3798	0.0039	0.010	0.0187	0.0009	0.049	0.1498	0.510
Total	0.3658	0.0096	0.026	0.0199	0.0013	0.065	0.1317*	0.602

Ks: synonymous substitution rate; Ka: nonsynonymous substitution rate; Ka/Ks: the ratio of nonsynonymous substitution rate to synonymous substitution rate; *α*: the shape parameter of the gamma distribution for rate heterogeneity; I: codon bias index; arithmetic: mean of *α* value for 13 genes.

To further study the selection pressure on the mitochondria of domestic horse, we surveyed the DNA sequence polymorphism of four genes (*ATP6, ATP8, ND6* and *CYTB*) in a larger size of sample. The four representatives were chosen because *ATP6* gene shows the highest level of diversity, *ND6* and *ATP8* genes exhibit higher substitution rates, and the Ka/Ks of *CYTB* gene is lower in the interspecies (horse *vs* donkey) and intraspecies comparisons (Tables **1** and **2**, Fig. **S2**).

Selective pressure may act differentially on genes within an organism. Here, non-directional selection was observed in *ND6*, *ATP6* and *CYTB* genes, but not *ATP8,* with a Tree-based relative ratio test by CRANN [51] (Fig. **3**). Moreover, selective pressure may act differentially on different domains and sites within genes; thus, we distinguished between surface and transmembrane domains (Table **S3**) as well as between synonymous and nonsynonymous sites in each of the four genes at the horse population level by neutrality tests (Table **3**). The surface of *ND6* gene and the transmembrane of the *CYTB* gene both indicate significantly negative values on the three neutrality tests. Interestingly, a significantly negative value of neutrality tests is found in the nonsynonymous sites of the four genes, except *ATP8*. It is striking that none of *ATP8* region has significant values for Tajima's *D*, Fu & Li's *D**, Fu & Li's *F**.

To scrutinize the potentially selected sites in the above four genes at a population level, tree topologies

Table 2. McDonald-Kreitman tests for mitochondrial protein-coding genes.

	Polymorphic[a]		Fixed		NI_{h-d}
	S	N	S	N	
ND1	17	5	57	4	4.191*
ND2	25	5	70	14	1.000
ND3	9	0	20	1	0
ND4L	6	2	18	2	3.000
ND4	39	18	95	6	7.308***
ND5	43	16	132	16	3.070**
ND6	7	6	22	8	2.357
COX1	42	6	95	1	13.571**
COX2	17	2	42	1	4.941
COX3	20	3	47	1	7.050
ATP6	22	7	46	6	2.439
ATP8	3	1	8	7	0.381
CYTB	29	6	75	2	7.759**
Total	279	77	727	69	2.908***

S: the number of synonymous substitution; N: the number of nonsynonymous substitution; NI: the neutral indices; h-d: comparison between horse and donkey; the p-value from Fisher's exact test (* < 0.5; ** < 0.1; *** < 0.01); superscript a: based on 20 horse mitochondrial genomes.

generated by CRANN were first modified with the maximum likelihood method by the PAML package. In the site-specific likelihood models, the likelihood ratio test (LRT) was performed by taking twice the difference of the log likelihood between nested models, and its result was compared with the value obtained by the chi-square distribution (Table **4**). The LRTs of M1 *vs* M2 and M7 *vs* M8 were not significant for *ND6* gene. Although two sites (118K and 162L) were all identified under positive selection by nested model comparison at more than 96% probability with BEB analysis, $2\Delta lnL_{M1vsM2} = 4.2336$ with $0.1 < p < 0.5$ and $2\Delta lnL_{M7vsM8} = 5.1531$ with $0.05 < p < 0.1$. One amino acid (12N) was found to have a low BEB probability ($p_{M1vsM2} = 0.753$ or $p_{M7vsM8} = 0.843$) for *ATP8* gene, and $2\Delta lnL$ of the two nested models were not significant in light of the chi-square distribution ($\chi^2 = 5.9915$, $df = 2$, $p = 0.05$). For the *ATP6* gene, the difference between M1 and M2 or M7 and M8 was not significant for $2\Delta lnL_{M1vsM2} = 2.9572$ ($0.1 < p < 0.5$) or $2\Delta lnL_{M7vsM8} = 3.8039$ ($0.1 < p < 0.5$), although two sites (63S and 15V) were detected as selected ones, with 92% probability for 63S. For the *CYTB* gene, the LRT of M1 *vs* M2 was not significant. Nevertheless, the difference between M7 and M8 was $2\Delta lnL = 7.2733$ with $0.01 < p < 0.05$ and $df = 2$, so M8 was significantly better than M7. Three amino acids, 214D ($p = 0.922$), 241T ($p = 0.620$) and 242L ($p = 0.601$), were distinguished as potentially selected sites with M7 *vs* M8. According to

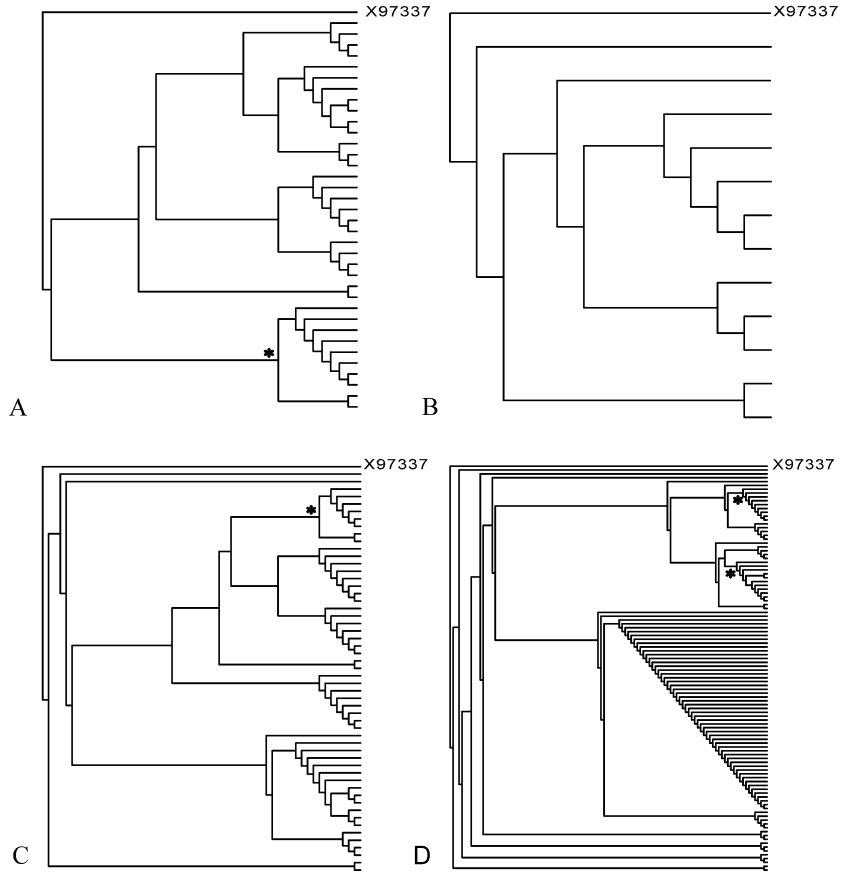

Fig. (3). Tree-based relative ratio test for non-directional selection. A: ND6; B: ATP8; C: ATP6; D: CYTB. Neighbor-joining trees under Ka/Ks distance; *: the selected branches.

Table 3. Neutrality tests on four mitochondrial protein-coding genes at population level.

	Region	Length (bp)	Tajima's *D*	Fu & Li's *D*	Fu & Li's *F*
ND6	CP	528	-1.813*	-2.769*	-2.896*
	SF	309	-2.296**	-3.263*	-3.463*
	TM	219	-0.635	-0.923	-0.976
	SY	135	-1.543	-2.009	-2.190
	NY	393	-1.898*	-3.041*	-3.145*
ATP8	CP	204	-1.428	-1.634	-1.798
	SF	150	-0.991	-1.037	-1.154
	TM	54	-1.094	-1.094	-1.11
	SY	49	-1.230	-1.320	-1.433
	NY	155	-1.713	-1.713	-1.919
ATP6	CP	681	-1.744	-3.098*	-3.096*
	SF	339	-1.761	-2.178	-2.340
	TM	342	-1.618	-2.340	-2.482
	SY	174	-1.542	-2.067	-2.236
	NY	507	-2.017*	-4.503**	-4.230**
CYTB	CP	1140	-1.690	-2.129	-2.337*
	SF	579	-1.418	-1.210	-1.548
	TM	561	-1.833*	-2.368*	-2.585*
	SY	286	-1.440	-1.789	-1.976
	NY	854	-2.182**	-2.444*	-2.799*

CP: complete gene; SF: the surface domain; TM: the transmembrane domain; SY: synonymous sites; NY: nonsynonymous sites; The p value (* < 0.5; ** < 0.1).

Table 4. Potentially selected amino acid sites for *ND6*, *ATP8*, *ATP6* and *CYTB* genes.

	PAML		ADAPTSITE
	M1 *vs* M2 (BEB)	**M7 *vs* M8 (BEB)**	**Maximum Likelihood**
ND6	$118K_s$*, $162L_t$*	$118K_s$**, $162L_t$**	$118K_s$***, $168V_t$***
ATP8	$12N_t$*	$12N_t$*	$12N_t$***
ATP6	$63S_s$*	$15V_t$*, $63S_s$*	$63S_s$***
CYTB	$214D_s$*	$214D_s$***, $241T_t$***, $242L_t$***	$23A_s$***, $156I_t$***, $209S_s$***, $214D_s$***, $242L_t$***, $254D_s$***, $316M_s$***

M1: nearly neutral model; M2: positive selection model; M7: beta model; M8: beta & ω model. Subscript s: the surface sites; Subscript t: the transmembrane sites. *0.1 < p < 0.5; ** 0.05 < p < 0.1; *** p < 0.05.

PAML manual, if the Ks sum of all branches on the tree is larger than 0.5, the LRTs will be reasonable. We summed Ks over all braches on the ML tree of *ND6* (Ks = 0.2248), *ATP8* (Ks = 0.1212), *ATP6* (Ks = 0.7591) and *CYTB* (Ks = 0.8866) genes. Hence, site 214D in the *CYTB* gene is the only one that had reliable evidence of selection according to results from LRT and the sum of Ks value.

Mitochondrial genes usually show a high rate of nucleotide substitution [22]. ADAPTSITE program was used because it is well suited to detect selection at highly polymorphic codons [67]. In the total, 11 sites were inferred as positively selected by the maximum likelihood approach accommodated within ADAPTSITE (version 1.6) ($p < 0.05$) (Table **4**). Evidence of positive selection in *ATP6*, *ND6* and *CYTB* genes is reliable because ADAPTSITE results are only dependable for aligned sequences with more than 15 nucleotide substitutions [67]. Although evolutionary models assumed in the PAML and ADAPTSITE software are different, five selected amino acid sites, 118K of the *ND6* gene, 12N of the *ATP8*, 63S of the *ATP6*, and 214D and 242L of the *CYTB* gene, were both detected by them.

Demographic History

To assess demographic scenarios of the domestic horse, we summarized the data as Tajima's *D*, and use

simulations to fit different demographic scenarios. This analysis could give us some clues about the ancestral population size changes in the past. Here, the observed Tajima's *D* is -1.3047, calculated from the polymorphism data of 20 horse mitochondrial genomes. This value is very similar to another Tajima's *D* (-1.3297) calculated from 363 concatenated (*ATP6*, *ATP8*, *ND6* and *CYTB genes*) sequences. So we take the *D* = -1.3 and compare it with simulated *D* values under four demographic models.

Based on the summary statistics, we were able to distinguish between different demographic scenarios. Fig. (**4**) suggests that the simulated Tajima's *D* values did not match the observed data (-1.3) under the bottleneck ($p < 0.01$), island ($p < 0.001$) and splitting ($p < 0.01$) models. The bottleneck severity does not significantly change the Tajima's *D* value given the bottleneck occurred 15,000-3,000 years before present (the most convincing domestication period of horses), partly because female lineages have been introduced into the domestic horse population over the period of domestication, and/or hybridization with wild horses in regions of horse domestication across the world [68-70]. The simulated Tajima's *D* value is slightly negative when the gene flow among subpopulations is high ($4Nm >= 5$). Moreover, the occurrence time did not

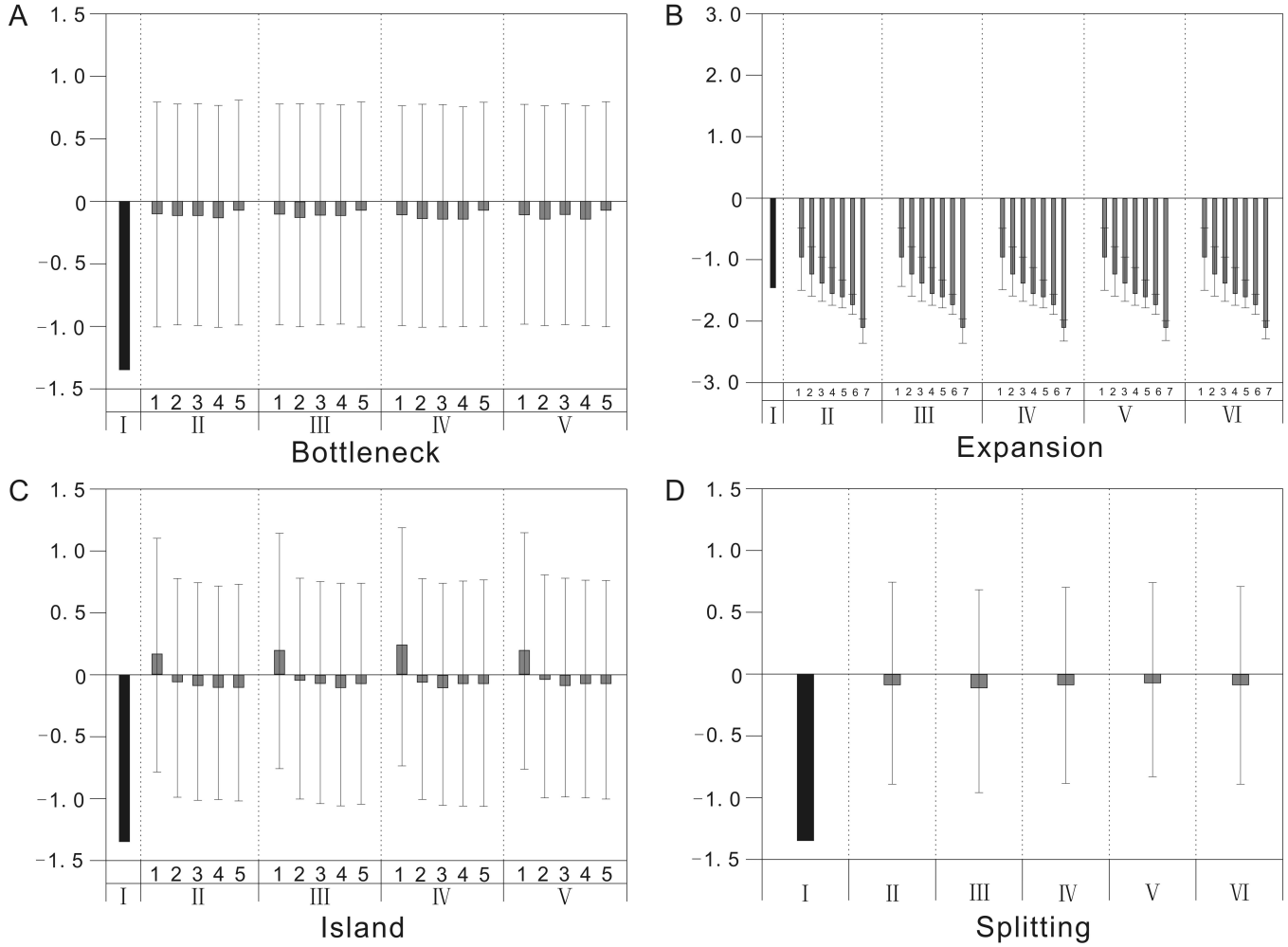

Fig. (4). Coalescent simulations of Tajima's *D* and its 95% confidence intervals under different population histories. A: Bottleneck model. 'I' represents the observed Tajima's *D*, while 'II, III, IV and V' represent the bottleneck occurred in 15,000, 10,000, 6,000 and 3,000 years ago, respectively. The number '1, 2, 3, 4 and 5' represent the simulated Tajima's *D* under the bottleneck severity $1/10^5$, $1/10^4$, $1/10^3$, $1/10^2$, 1/10, respectively. **B**: Expansion model. 'I' represents the observed Tajima's *D*, while 'II, III, IV, V and VI' represent the expansion occurred in 1,000, 2,000, 3,000, 4,000 and 5,000 years ago, respectively. The number '1, 2, 3, 4, 5, 6 and 7' represent the simulated Tajima's *D* under the expansion severity 10, 20, 30, 40, 50, 100 and 1,000, respectively. **C**: Island model. 'I' represents the observed Tajima's *D*, while 'II, III, IV and V' represent the subpopulation situation: 2 subpopulations with a size ratio 1:1, 3 subpopulations with a size ratio 2:1:1, 4 subpopulations with a size ratio 1:1:1:1, 4 subpopulations with a size ratio 4:3:2:1, respectively. The number '1, 2, 3, 4 and 5' represent the simulated Tajima's *D* under the level of gene flow 1, 5, 10, 100 and 1,000 between subpopulations, respectively. **D**: Splitting model. 'I' represents the observed Tajima's *D*, while 'II, III, IV, V and VI' represent the split occurred in 1,000, 2,000, 3,000, 4,000 and 5,000 years ago, respectively. The main population was split into 2 equal size subpopulations without gene flow at each split time. Light grey columns represent simulated Tajima's *D* of the split time 1,000, 2,000, 3,000, 4,000 and 5,000 years ago, respectively.

obviously change the Tajima's *D* under the two-equal-splitting subpopulation model without gene flow after horse domestication 5,500 years ago. It is most likely that the population expansion scenario is the most plausible matrilineal history for the domestic horse, and population sizes increased about 20-50 times.

Multiple Hits

Under the infinite site model [71], only a single mutation can occur at one site. But this assumption is very restricted and can be invalid here because of the high mutation rate of mitochondria. If so, multiple hits or recurrent mutations could happen. We did find 1199 multiple-hit sites in the 20 horse mitochondrial genomes analyzed and 251 sites in the concatenated horse sequences (*ATP6*, *ATP8*, *ND6* and *CYTB* genes).

DISCUSSION

Substitution Rate Heterogeneity of Mitochondria

The substitution rate across the mitochondrial genome of the domestic horse is highly heterogeneous, demonstrated by the relative nucleotide substitution rates (Ka/Ks) and the shape parameter α for 13 mitochondrial protein-coding genes (Table **1**). Our study indicates that the horse mitochondrial genome has a similar mutation rate and selective constraint at synonymous sites, however, we observed a substitution rate heterogeneity and different selective constraints on nonsynonymous sites. The pattern of heterogeneity on nonsynonymous sites is similar to that seen in previous studies of *Drosophila* and man [23, 72], but differs from a study in gadine fish, where the substitution rate heterogeneity was documented for both synonymous and nonsynonymous sites [73].

The heterogeneity for the nonsynonymous substitution rate is possibly due in part to different selective pressures among genes. Although mutation and/or selection may influence evolution rate, mutation causes proportional variations to both synonymous and nonsynonymous substitutions, while positive selection results in an increased rate of nonsynonymous substitutions relative to synonymous substitutions [74]. In our study, synonymous substitutions vary slightly, whereas nonsynonymous substitutions change markedly. Genes such as the *ATP8* and *ND6* exhibit an elevated nonsynonymous substitution rate but a decreased synonymous substitution rate. Moreover, the Ka for 13 protein-coding genes in our study does not show a significantly positive correlation with the distance (Dss) from the light-strand origin of replication ($r = 0.219$, $p = 0.472$). However, when the four exceptional (*ATP6*, *ATP8*, *ND3* and *CYTB*; | studentized residuals | > 1.4) genes are omitted, a strong positive relationship is observed between the Ka and the gene's position ($r = 0.749$, $p = 0.02$). A previous study documented that spatial heterogeneity of nonsynonymous substitution rates shows a significantly positive correlation with variation in mutation rates, which is related to the mtDNA replication mechanism [75].

Pattern of the Horse Mitochondrial Genome Evolution Shaped by Selection

Mitochondria play an important role in energy metabolism and oxygen usage, suggesting that mitochondrial genes may evolve under functional constraints. However, an increasing number of cases of selection on mitochondrial protein-coding genes have been detected in wild animals, such as pike, antelope, South American camelids and Chinese sub-nosed monkey [6, 14, 76-79]. In this study, the mitochondrial protein-coding genes of domestic horse were analyzed at both mitochondrial genome level and population level, to address the role of positive selection in the evolution of mtDNA of domestic horse.

Signatures of positive selection were found in mtDNA of domestic horse. Here, non-directional selection (variation for maintaining an acceptable level of fitness to changes in the environment) was identified for *ND6*, *ATP6* and *CYTB* genes (Fig. **3**). Moreover, the signature of positive selection was detected on the surface and the transmembrane domain of the *ND6* and the *CYTB* genes, respectively. When we did the site-specific tests for positive selection, five amino acid sites in *ND6*, *ATP6*, *ATP8* and *CYTB* genes were identified by both PAML and ADAPTSITE softwares (Table **4**).

The possible cause for positive selection on each of these genes is related to genes' function. The reason for why are some genes under selection is due to their different roles in the process of OXPHOS. The OXPHOS system is composed of five complexes: (i) NADH-CoQ oxidoreductase, (ii) succinate dehydrogenase, (iii cytochrome bc1, (iv) cytochrome c oxidase and (v) ATP synthase. MtDNAs encode seven protein subunits (mitochondrial subunits I: *ND1-6*, *ND4L*) of OXPHOS complex I, one subunit (mitochondrial subunits II: *cyochrome b*) of OXPHOS III, three subunits (mitochondrial subunits III: *COI-III*) of OXPHOS complex IV and two subunits (mitochondrial subunits IV: *ATP6, 8*) of OXPHOS complex V. The mitochondrial electron transport system performs the process of OXPHOS. Electrons enter the chain at either complexes I or II and are transported between complexes by the carrier molecules ubiquinone and coenzyme. This process creates potential energy in the form of protons (H^+), which are pumped from the complexes into the intermembrane space. Electrons are delivered to oxygen by complex IV. Finally, the protons are recruited by complex V (ATP synthase), which acts as a molecular motor to convert the potential energy into ATP. Mitochondrial subunits I and II perform the known catalytic functions COX and are major sources of reactive oxygen species, which are potent mutagens [80]. There is evidence that the cytochrome b (*CYTB*) has an accelerated non-synonymous substitution rate in anthropoid primates [81]. In this study, we found positive selection on mitochondrial *CYTB* gene in the domestic horse. Moreover, evidence of positive selection was found in the *ND2* and *ND6* genes of Chinese snub-nosed monkeys [79]. Our previous study also demonstrated

adaptive evolution of the mitochondrial *ND6* gene in the domestic horse [39]. Highly frequent genetic variations are found in horse mtDNA in this study. Thus, positive selection might be due to standing genetic variation instead of newly derived alleles. In brief, function difference cause various representation of the 13 mitochondrial-encoded genes under selection pressure.

Although only four mitochondrial genes were analyzed at a population level, our study shed light on the influence of positive selection on mitochondrial genome evolution in domestic horse. Analyses of structural and functional changes of mitochondrial genomes in domestic horse, as well as more mitochondrial genome data from the domestic horse, are needed for reveal the exact molecular mechanisms of positive selection in the species.

Multiple Hits, Rate Heterogeneity and False Positives for Recombination

The role of recombination within mitochondrial genomes remains a topic of considerable debate. Usually, recombination of mtDNA is thought to be absent in animals [82]. Paternal inheritance of mtDNA has been found in *Drosophila*, great tit, mice, honeybee, cattle and human [29, 30, 83-87], indicating that recombination is possible in the mtDNA of these species. Direct evidence for recombination in mitochondria is documented in a few species, such as nematode, flatfish, mussel and human [88-92], and indirect evidence exists in more animal species [31, 32]. However, many evolutionary factors can lead to false positive evidence for recombination identified by indirect tests [93]. Here, we discuss whether recombination in horse mtDNA could be falsely inferred due to multiple hits and rate heterogeneity.

Multiple hits may cause similar patterns raised by recombination (*e.g.* 4-gamete [64]). We found 1199 multiple-hit sites in 20 complete horse mitochondrial genomes, and 251 multiple-hit sites in concatenated sequences of *ATP6*, *ATP8*, *ND6* and *CYTB* genes. The large number of multiple hits indicates a high mutation rate under a finite site mutation model, which agrees that multiple hits account for the false positives of recombination [65, 94].

Rate heterogeneity could be another factor leading to false positives for recombination [93, 95]. In this study, 13 mitochondrial protein-coding genes possess different shape parameters (α) of the gamma distribution which has been used to model the variation in mutation rate [96-99]. Moreover, nonsynonymous substitutions exhibited strong rate heterogeneity in horse mtDNA. Rate heterogeneity of mutation and substitution in horse mtDNA may lead to false positives for recombination.

Considering the high number of multiple hits and the strong rate heterogeneity observed in horse mtDNA sequences, it would be extremely difficult to distinguish the effect of recombination from them. There is no strong evidence of recombination of horse mtDNA, although recombination has been indicated by indirect tests (Table **S4**).

ABBREVIATIONS

α	=	The shape parameter of the gamma distribution for rate heterogeneity
π	=	Nucleotide diversity
BEB	=	Bayes empirical Bayes
I	=	Codon bias index
Ka	=	Nonsynonymous substitution rate
Ka/Ks	=	The ratio of nonsynonymous substitution rate to synonymous substitution rate
Ks	=	The synonymous substitution rate
mtDNA	=	Mitochondrial DNA
N	=	The number of nonsynonymous nucleotide substitution
np	=	Nucleotide position
nt	=	Nucleotides
Numts	=	Nuclear mitochondrial pseudogenes
OXPHOS	=	Oxidative phosphorylation
S	=	The number of synonymous nucleotide substitution

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We thank G.D. Wang and H.B. Xie for support prior to and during the course of data analyses. S.F. Wu and T. Sha are gratefully acknowledged for sample collection. This work was funded by grants provided by the National Program on Key Basic Research Project China (2007CB815700), the Science Foundation of the Chinese Academy of Sciences, the NSFC and Yunnan Province Union Fund (U0836602), the National Natural Science Foundation of China (30760038，31172073 and 31301014), the Applied Basic Research Programs of Yunnan Province (2011FZ001), the Scientific Research Foundation of Education Bureau of Yunnan Province (2010Y502 and 2011Y104) and by the Scientific Research Foundation of Kunming University (YJL11003).

SUPPLEMENTARY MATERIALS

Supplementary material is available on the publisher's web site along with the published article.

Table **S1**. Sampling information of 509 domestic horses collected across mainland China.

Table **S2**. Accession numbers of mitochondrial genome sequences of large draft livestocks that were analyzed in this study.

Table **S3**. Surface and transmembrane domains of *ND6*, *ATP8*, *ATP6* and *CYTB* genes.

Table **S4**. List of six indirect recombination tests on 20 horse mitochondrial genomes.

Fig. **S1**. Schematic diagrams of demographic models (see Fig. **4** as well as MATERIALS AND METHODS section for details). N: effective population size; T: the time in years measured from the present

Fig. **S2**. Nucleotide diversity of mitochondrial genes. The mean diversity is indicated by the broken line.

Fig. **S3**. Nucleotide diversity of mitochondrial genomes among large draft livestocks.

REFERENCES

[1] Zeder MA. Central questions in the domestication of plants and animals. Evol Anthropol 2006; 15: 105-17.

[2] Outram AK, Stear NA, Bendrey R, *et al*. The earliest horse harnessing and milking. Science 2009; 323: 1332-5.

[3] Lopez-Barneo J, Pardal R, Ortega-Saenz P. Cellular mechanism of oxygen sensing. Annu Rev Physiol 2001; 63: 259-87.

[4] Grossman LI, Wildman DE, Schmidt TR, Goodman M. Accelerated evolution of the electron transport chain in anthropoid primates. Trens Genet 2004; 20: 578-85.

[5] Blier PU, Dufresne F, Burton RS. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. Trends Genet 2001; 17: 400-6.

[6] da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics 2008; 9: 119.

[7] Rand DM, Haney RA, Fry AJ. Cytonuclear coevolution: the genomics of cooperation. Trends Ecol Evol 2004; 19: 645-53.

[8] Bayona-Bafaluy MP, Müller S, Moraes CT. Fast adaptive coevolution of nuclear and mitochondrial subunits of ATP synthetase in orangutan. Mol Biol Evol 2005; 22: 716-24.

[9] Edmands S, Burton RS. Cytochrome c oxidase activity in interpopulation hybrids of a marine copepod: A test for nuclear-nuclear or nuclear-cytoplasmic coadaptation. Evolution 1999; 53:1972-8.

[10] Willett CS, Burton RS. Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod. Mol Biol Evol 2004; 21:443-53.

[11] Parmakelis A, Kotsakiozi P, Rand D. Animal mitochondria, positive selection and cytonuclear coevolution: insights from pulmonates. PLoS One 2013; 8: e61970.

[12] Gershoni M, Fuchs A, Shani N, *et al*. Coevolution predicts direct interactions between mtDNA-encoded and nDNA-encoded subunits of oxidative phosphorylation complex i. J Mol Biol 2010; 19: 158-71.

[13] Rand DM. The units of selection on mitochondrial DNA. Annu Rev Ecol Syst 2001; 32: 415-48.

[14] Hassanin A, Ropiquet A, Couloux A, Cruaud C. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). J Mol Evol 2009; 68: 293-310.

[15] Jobson RW, Nielsen R, Laakkonen L, Wikstrom M, Albert VA. Adaptive evolution of cytochrome c oxidase: Infrastructure for a carnivorous plant radiation. Proc Natl Acad Sci U S A 2004; 101: 18064-8.

[16] Dimauro S, Davidzon G. Mitochondrial DNA and disease. Ann Med 2005; 37: 222-32.

[17] Johns DR. The other human genome: mitochondrial DNA and disease. Nat Med 1996; 2: 1065-8.

[18] Wallace DC. Mitochondrial diseases in man and mouse. Science 1999; 283: 1482-8.

[19] Shoffner JM. Molecular analysis of oxidative phosphorylation diseases for detection of mitochondrial DNA mutations. Curr Protoc Hum Genet 2001; Chapter9: Unit 9.9.

[20] Nachman MW, Boyer SN, Aquadro CF. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proc Natl Acad Sci U S A 1994; 91: 6364-8.

[21] Moilanen JS, Finnila S, Majamaa K. Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J. Mol Biol Evol 2003; 20: 2132-42.

[22] Ballard JW, Whitlock MC. The incomplete natural history of mitochondria. Mol Ecol 2004; 13: 729-44.

[23] Elson JL, Turnbull DM, Howell N. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am J Hum Genet 2004; 74: 229-38.

[24] Moyer GR, Winemiller KO, McPhee MV, Turner TF. Historical demography, selection, and coalescence of mitochondrial and nuclear (genes in Prochilodus species of northern South America. Evolution 2005; 59: 599-610.

[25] Weinreich DM, Rand DM. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. Genetics 2000; 156: 385-99.

[26] Gerber AS, Loggins R, Kumar S, Dowling TE. Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? Annu Rev Genet 2001; 35: 539-66.

[27] Skibinski DO, Gallagher C, Beynon CM. Mitochondrial DNA inheritance. Nature 1994; 368: 817-8.

[28] Zouros E, Ball AO, Saavedra C, Freeman KR. Mitochondrial DNA inheritance. Nature 1994; 368: 818.

[29] Kvist L, Martens J, Nazarenko AA, Orell M. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). Mol Biol Evol 2003; 20: 243-7.

[30] Zhao X, Li N, Guo W, *et al*. Further evidence for paternal inheritance of mitochondrial DNA in the sheep (*Ovis aries*). Heredity 2004; 93:399-403.

[31] Piganeau G, Gardner M, Eyre-Walker A. A broad survey of recombination in animal mitochondria. Mol Biol Evol 2004; 21: 2319-25.

[32] Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E. Widespread recombination in published animal mtDNA sequences. Mol Biol Evol 2005; 22: 925-33.

[33] D'Aurelio M, Gajewski CD, Lin MT, *et al*. Heterologous mitochondrial DNA recombination in human cells. Hum Mol Genet 2004; 13: 3171-9.

[34] Hoolahan AH, Blok VC, Gibson T, *et al*. Evidence of animal mtDNA recombination between divergent populations of the potato cyst nematode *Globodera pallida*. Genetica 2012; 140: 19-29.

[35] Peters J, von den Driesch A, Helmer D. The First Steps of Animal Domestication. Oxford: Oxbow Books 2005.

[36] Bjornerfeldt S, Webster MT, Vilà C. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Res 2006; 16: 990-4.

[37] Wang Z, Yonezawa T, Liu B, *et al*. Domestication relaxed selective constraints on the yak mitochondrial genome. Mol Biol Evol 2011; 28: 1553-6.

[38] Xu S, Luosang J, Hua S, *et al*. High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. J Genet Genomics 2007; 34: 720-9.

[39] Ning T, Xiao H, Li J, Hua S, Zhang YP. Adaptive evolution of the mitochondrial ND6 gene in the domestic horse. Genet Mol Res 2010; 9: 144-50.

[40] China National Commission of Animal Genetic Resource. Animal Genetic Resource in China, Horses Donkeys Camels. Beijing: China Agriculture Press 2011.

[41] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22: 4673-80.

[42] Xu X, Arnason U. The complete mitochondrial DNA sequence of the horse, Equus caballus: extensive heteroplasmy of the control region. Gene 1994; 148: 357-62.

[43] Xia X, Xie Z. DAMBE: software package for data analysis in molecular biology and evolution. J Hered 2001; 92: 371-3.

[44] Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009; 25: 1451-2.
[45] Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 2006; 4: 259-63.
[46] Tajima F. The effect of change in population size on DNA polymorphism. Genetics 1989; 123: 597-601.
[47] Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics 1993; 133: 693-709.
[48] Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 1997; 147: 915-25.
[49] McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature 1991; 351: 652-4.
[50] Hofmann K, Stoffel W. TMbase-—a database of membrane spanning proteins segments. J Biol Chem 1993; 374: 166.
[51] Creevey CJ, McInerney JO. An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. Gene 2002; 300: 43-51.
[52] Yang Z. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J Mol Evol 2000; 51: 423-32.
[53] Suzuki Y, Gojobori T, Nei M. ADAPTSITE: detecting natural selection at single amino acid sites. Bioinformatics 2001; 17: 660-1.
[54] Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 2005; 22: 1107-18.
[55] Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 2002; 18: 337-8.
[56] Wright S. Coefficients of inbreeding and relationship. American Naturalist 1922; 56: 330-338.
[57] Mills DS, McDonnell SM. The Domestic Horse: The Origins, Development and Management of its Behaviour. Cambridge: Cambridge University Press 2005.
[58] Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. Bioinformatics 2000; 16: 562-3.
[59] Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. Virology 1999; 265: 218-25.
[60] Martin DP, Posada D, Crandall KA, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retroviruses 2005; 21: 98-102.
[61] Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 2000; 16: 573-82.
[62] Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci U S A 2001; 98: 13757-62.
[63] Martin DP, Williamson C, Posada D. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 2005; 21: 260-2.
[64] McVean GA. What do patterns of genetic variability reveal about mitochondrial recombination? Heredity 2001; 87: 613-20.
[65] Fitch WM. Toward Defining Course of Evolution: Minimum Change for a Specified Tree Topology Syst Zool 1971; 20: 406-16.
[66] Zink RM, Drovetski SV, Rohwer S. Selective neutrality of mitochondrial ND2 sequences, phylogeography and species limits in Sitta europaea. Mol Phylogenet Evol 2006; 40: 679-86.
[67] Suzuki Y. New methods for detecting positive selection at single amino acid sites. J Mol Evol 2004; 59: 11-9.
[68] Vila C, Leonard JA, Gotherstrom A, *et al*. Widespread origins of domestic horse lineages. Science 2001; 291: 474-7.
[69] Jansen T, Forster P, Levine MA, *et al*. Mitochondrial DNA and the origins of the domestic horse. Proc Natl Acad Sci U S A 2002; 99: 10905-10.
[70] Lippold S, Matzke NJ, Reissmann M, Hofreiter M. Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. BMC Evol Biol 2011; 11: 328.
[71] Kimura M. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 1969; 61: 893-903.
[72] Ballard JW. Comparative genomics of mitochondrial DNA in Drosophila simulans. J Mol Evol 2000; 51: 64-75.
[73] Marshall HD, Coulson MW, Carr SM. Near neutrality, rate heterogeneity, and linkage govern mitochondrial genome evolution in Atlantic cod (*Gadus morhua*) and other gadine fish. Mol Biol Evol 2009; 26: 579-89.
[74] Kreitman M, Akashi H. Molecular evidence for natural selection. Annual Review of Ecology and Systematics 1995; 26: 403-22.
[75] Broughton RE, Reneau PC. Spatial covariation of mutation and nonsynonymous substitution rates in vertebrate mitochondrial genomes. Mol Biol Evol 2006; 23: 1516-24.
[76] Xu SQ, Yang YZ, Zhou J, *et al*. A mitochondrial genome sequence of the Tibetan antelope (*Pantholops hodgsonii*). Genomics Proteomics Bioinformatics 2005; 3: 5-17.
[77] Luo Y, Gao W, Gao Y, *et al*. Mitochondrial genome analysis of Ochotona curzoniae and implication of cytochrome c oxidase in hypoxic adaptation. Mitochondrion 2008; 8: 352-7.
[78] Di Rocco F, Zambelli AD, Vidal Rioja LB. Identification of camelid specific residues in mitochondrial ATP synthase subunits. J Bioenerg Biomembr 2009; 41: 223-8.
[79] Yu L, Wang X, Ting N, Zhang Y. Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. Mitochondrion 2011; 11: 497-503.
[80] Finkel T, Holbrook NJ. Oxidants, oxidative stress and the biology of aging. Nature 2000; 408: 239-47.
[81] Andrews TD, Jermiin LS, Easteal S. Accelerated evolution of cytochrome b in simian primates: adaptive evolution in concert with other mitochondrial proteins? J Mol Evol 1998; 47: 249-257.
[82] Avise JC. Molecular markers, natural history and evolution. London: Chapman and Hall 1994.
[83] Kondo R, Satta Y, Matsuura ET, Ishiwa H, Takahata N, Chigusa SI. Incomplete maternal transmission of mitochondrial DNA in Drosophila. Genetics 1990; 126: 657-63.
[84] Gyllensten U, Wharton D, Josefsson A, Wilson AC. Paternal inheritance of mitochondrial DNA in mice. Nature 1991; 352: 255-7.
[85] Meusel MS, Moritz RF. Transfer of paternal mitochondrial DNA during fertilization of honeybee (*Apis mellifera* L.) eggs. Curr Genet 1993; 24: 539-43.
[86] Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G. Ubiquitin tag for sperm mitochondria. Nature 1999; 402: 371-2.
[87] Schwartz M, Vissing J. Paternal inheritance of mitochondrial DNA. N Engl J Med 2002; 347: 576-80.
[88] Lunt DH, Hyman BC. Animal mitochondrial DNA recombination. Nature 1997; 387: 247.
[89] Ladoukakis ED, Zouros E. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. Mol Biol Evol 2001; 18: 1168-75.
[90] Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish Platichthys flesus. Mol Biol Evol 2002; 19: 2261-4.
[91] Burzynski A, Zbawicka M, Skibinski DO, Wenne R. Evidence for recombination of mtDNA in the marine mussel Mytilus trossulus from the Baltic. Mol Biol Evol 2003; 20: 388-92.
[92] Kraytsberg Y, Schwartz M, Brown TA, *et al*. Recombination of human mitochondrial DNA. Science 2004; 304: 981.
[93] Sun S, Evans BJ, Golding GB. "Patchy-tachy" leads to false positives for recombination. Mol Biol Evol 2011; 28: 2549-59.
[94] Innan H, Nordborg M. Recombination or mutational hot spots in human mtDNA? Mol Biol Evol 2002; 19: 1122-7.

[95] Hagelberg E. Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. Trends Genet 2003; 19: 84-90.

[96] Nei M, Chakraborty R, Fuerst PA. Infinite allele model with varying mutation rate. Proc Natl Acad Sci U S A 1976; 73: 4164-8.

[97] Gu X, Fu YX, Li WH. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol Biol Evol 1995; 12: 546-57.

[98] Yang Z, Kumar S. Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. Mol Biol Evol 1996; 13: 650-9.

[99] Tourasse NJ, Gouy M. Evolutionary distances between nucleotide sequences based on the distribution of substitution rates among sites as estimated by parsimony. Mol Biol Evol 1997; 14: 287-98.

Received: November 26, 2013 Revised: April 14, 2014 Accepted: April 17, 2014