



## Adaptive evolution of the mitochondrial *ND6* gene in the domestic horse

T. Ning<sup>1,2</sup>, H. Xiao<sup>3</sup>, J. Li<sup>1,2</sup>, S. Hua<sup>3</sup> and Y.P. Zhang<sup>1,2</sup>

<sup>1</sup>Laboratory for Conservation and Utilization of Bio-Resource, Yunnan University, Kunming, Yunnan, China

<sup>2</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

<sup>3</sup>School of Life Sciences, Yunnan University, Kunming, Yunnan, China

Corresponding author: Y.P. Zhang

E-mail: zhangyp@mail.kiz.ac.cn

Genet. Mol. Res. 9 (1): 144-150 (2010)

Received October 1, 2009

Accepted November 4, 2009

Published January 26, 2010

**ABSTRACT.** Mitochondria play a crucial role in energy metabolism through oxidative phosphorylation. Organisms living at high altitudes are potentially influenced by oxygen deficits and cold temperatures. The severe environmental conditions can impact on metabolism and direct selection of mitochondrial DNA. As a wide-ranging animal, the domestic horse (*Equus caballus*) has developed various morphological and physiological characteristics for adapting to different altitudes. Thus, this is a good species for studying adaption to high altitudes at a molecular level. We sequenced the complete NADH dehydrogenase 6 gene (*ND6*) of 509 horses from 24 sampling locations. By comparative analysis of three horse populations living at different altitudes (>2200 m, 1200-1700 m, and <900 m), we found that the high-altitude population had the lowest genetic diversity and significant negative Tajima's *D*; both values declined with increasing elevation. Moreover, non-directional selection was identified for the *ND6* gene by a tree-based relative ratio test ( $P = 0.007$ ); the highest proportion of high-altitude horses was found distributed on the selected branches. We conclude that the high-altitude environment has directed adaptive evolution of the mitochondrial *ND6* gene in the plateau horse.

**Key words:** Domestic horse; *ND6*; Adaptive evolution

## INTRODUCTION

The mitochondrion is a key organelle involved in metabolism. It produces 95% of a eukaryotic cell's energy and encodes 13 proteins that compose four complexes in mammals: NADH dehydrogenase (*ND1*, 2, 3, 4, 4L, 5, and 6), cytochrome bc1 (*CytB*), cytochrome c oxidase (*COXI*, 2, 3), and ATP synthase (*ATP6* and *ATP8*). All of these four complexes involved in the oxidative phosphorylation process are essential for energy metabolism (Lopez-Barneo et al., 2001). Hypoxia and cold are two characteristics of a plateau environment; they can directly influence metabolic performance and modify selective pressures on mitochondria. Several investigators have reported physiological and genetic (including hemoglobin and mitochondrial gene) adaptations to high-altitude environments in humans, deer, mice, and cattle Caparini (Beall, 2003; Weber, 2007; Gering et al., 2009; Hassanin et al., 2009). Although physiological adaptation of domestic horses living at high altitudes has also been reported (Wickler and Greene, 2003), only one study examined the responses of mitochondrial DNA to this environment in the domestic horse. The *ND6* gene was reported to accumulate larger numbers of nonsynonymous substitutions, based on analysis of three mitochondrial genomes of the Tibetan plateau horse, suggesting high-altitude adaptations at the *ND6* gene in domestic horses (Xu et al., 2007). However, the dataset was small and there was no comparative analysis between plateau and plain horses. We investigated whether the *ND6* gene has been affected by adaptive evolution in the plateau (high altitude) horse population.

## MATERIAL AND METHODS

### Sampling and sequencing

The complete *ND6* gene (528 bp) was amplified and sequenced for 509 Chinese domestic horses from 24 sampling locations using primers N6F (5'-CCAAAATCTATCTC CCAGTT-3') and N6R (5'-GAGCCGATTTCATCAT-3'). All *ND6* sequences in this study were evaluated by bi-directional sequencing and deposited as haplotypes in GenBank under accession numbers FJ643640-FJ643675.

### Population genetics analysis

For comparative analysis, all 509 Chinese horse samples were classified into three populations according to their elevation environments (>2200 m; 1200-1700 m; <900 m; Supplementary Table 1 and Supplementary Figure 1). Arlequin 3.0 (Excoffier et al., 2005) was used to calculate nucleotide and haplotype diversity and migrant number ( $N_m$ ) in these populations (Hudson et al., 1992). Tajima's  $D$  (Tajima, 1989) was calculated using DnaSP 4.0 (Rozas et al., 2003). Statistical significance of the Tajima's  $D$  test was assessed by  $10^4$  coalescent simulations without recombination.

### Detecting selective signal

To detect selection on the *ND6* gene in the horses, we first adopted a tree-based

relative ratio test to distinguish directional selection (successive amino acid change for efficiently performing a particular task) and non-directional selection (variation for maintaining an acceptable level of fitness to changes in the environment) by CRANN (Creevey and McInerney, 2002). A domestic donkey sequence (X97337) was used as the outgroup. The expectation of this test is that the ratio of replacement to silent polymorphisms on each branch should be equal to the ratio of replacement to silent substitutions over the entire tree under neutrality. The G test and the Fisher exact test were used to check significant differences in this ratio. The site-specific likelihood test with the CODEML module in the PAML package (Yang, 1997) was used to identify the potential selected sites of the *ND6* gene, discriminated by Bayes empirical Bayes analysis. Two nested models, M1 (nearly neutral) vs M2 (positive selection), and M7 ( $\beta$ ) vs M8 ( $\beta$  and  $\omega$ ) were used for the likelihood ratio tests. The significance of twice the difference in maximum likelihood values between nested models was checked with the chi-square distribution.

## RESULTS AND DISCUSSION

### Potential selected sites and effects of selection on genetic diversity

Thirty-six haplotypes were defined by 509 Chinese domestic horse sequences (Table 1). In general, if selection is operating on an organism, advantageous mutations will increase in frequency to adapt; after they have spread through a population, the level of diversity will be reduced. When we pooled 509 horse samples into three populations based on altitudes (>2200 m; 1200-1700 m; <900 m), the highest nucleotide diversity was found at the lowest altitude ( $\pi = 0.005$ ) and the lowest one at the highest altitude ( $\pi = 0.003$ ). Haplotype diversity was 0.6242, 0.7151 and 0.7936 at high, mid and low altitudes, respectively. Notably, both nucleotide and haplotype diversities of the *ND6* gene declined with increasing altitude (Table 1). Also the sites (118 K and 162 L) at the *ND6* gene were identified with more than 96% probability under positive selection by comparing two nested models in PAML (Table 2). A negative value of Tajima's *D*, which may reflect an effect of positive selection (Nielsen, 2005), was detected in all three altitude populations; it also declined with increasing altitude, but significance was only found for >2200 m and 1200-1700 m, after  $10^4$  coalescence simulations.

**Table 1.** Genetic statistics for the mitochondrial *ND6* gene in Chinese domestic horses

Po	N	n	h (SD)	$\pi$ (SD)	<i>D</i>	Nm		
						HP	MP	LP
All	509	36	0.7097 (0.0190)	0.0041 (0.0002)	-1.5452*			
HP	198	17	0.6242 (0.0354)	0.0030 (0.0020)	-1.5473*			
MP	240	26	0.7251 (0.0270)	0.0045 (0.0027)	-1.5433*	17.6986		
LP	71	11	0.7936 (0.0239)	0.0050 (0.0030)	-0.7818	2.8086	10.9338	

Po, N, n, h,  $\pi$ , *D*, and Nm denote the population, sample size, number of haplotypes, haplotype diversity, nucleotide diversity, Tajima's *D*, and migrant number, respectively. Nominal P values of the Tajima's *D* test were simulated with  $10^4$  non-recombination coalescence. Four Chinese domestic horse populations: All = a total of 509 samples; HP = altitude >2200 m; MP = altitude 1200-1700 m; LP = altitude <900 m. \*P < 0.05.

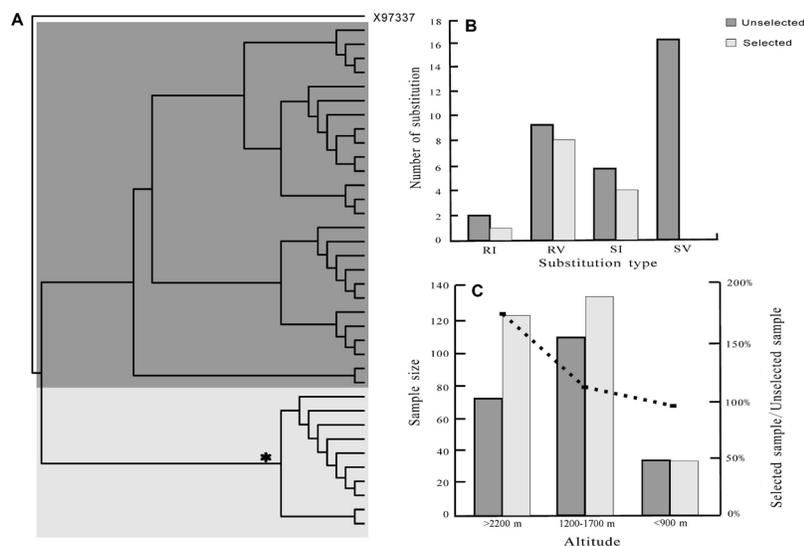
**Table 2.** Likelihood values and potential selected sites for the mitochondrial *ND6* gene.

Nested-models	Potential selected sites (BEB P)	d.f.	2 $\Delta$ lnL (P)
M1 vs M2	118 K (P = 0.995), 162 L (P = 0.967)	2	4.2336 (0.1 < P < 0.5)
M7 vs M8	118 K (P = 0.995), 162 L (P = 0.963)	2	5.1531 (0.05 < P < 0.1)

M1 = nearly neutral model; M2 = positive selection model; M7 =  $\beta$  model; M8 =  $\beta$  and  $\omega$  model. BEB = Bayes empirical Bayes analysis. d.f. = degrees of freedom between the respective models.

### Non-directional selection on the *ND6* gene

In the tree-based relative ratio test, non-directional selection branches ( $P = 0.007$ ) were identified in the neighbor-joining tree of the *ND6* gene constructed with Ka/Ks distances (Figure 1A). When we compared the sample distribution on the selected branches with that on the unselected branches in the three populations, we found three noteworthy results. First, the silent polymorphisms, which should be numerous in a population, were not detected on the selected branches (Figure 1B). This is expected if selective pressure has acted on the horses belonging to the selected branches and deleted silent polymorphism substitutions. Second, the distribution ratio of the selected to unselected horses increased from 97% (at the lowest latitude) to 122% (at the middle latitude) to 175% (at the highest latitude; Figure 1C). Third, the number of >2200 m horses and of <900 m horses are significantly different between the selected and unselected branches (chi-square test:  $P = 0.03$ ).



**Figure 1.** Non-directional selection and substitutions. The selected and unselected are marked with dark and light grey color, respectively. **A.** The neighbor-joining tree constructed with Ka/Ks distances by CRANN. \*Indicates branches that underwent non-directional selection. **B.** Number of substitutions distributed on the selected and unselected branches. RV = the replacement polymorphisms; SV = the silent polymorphisms; RI = the replacement substitutions; SI = the silent substitutions. **C.** Sample size and the distribution ratio on the selected versus unselected branches in the three populations. Small black rectangles connected with a broken line show the distribution ratio of samples on the selected compared to unselected branches.

### What accounts for the high proportion but not all plateau horses on selected branches?

Four factors help to explain this scenario. First, the  $N_m$  values show gene flow between the three horse populations. The largest value ( $N_m = 17.699$ ) was found between the high- and mid-altitude populations. This means that the three horse populations are not independent of each other. Second, several physiological studies provide considerable empirical evidence to support different individual fitness at high-altitude environments (Monge and Leon-Velarde, 1991; Schmidt-Nielsen, 1997; Wickler and Greene, 2003). Thus, individual fitness should be different on a molecular level within a population. Third, the earliest known domestication of horses was traced back to 3500 B.C. in the Eneolithic Botai Culture of Kazakhstan (Outram et al., 2009). Based on the short demographic history, evidence for selection (especially non-directional selection) would not likely be fixed on the mitochondrial DNA of all high-altitude horses. Fourth, recombination rarely occurs in mitochondria, so that if a substitution is under positive selection then any other substitutions in that lineage will be carried with it, which makes the infrequent adaptive evolution signature even more easily obscured by the ubiquitous negative selection within a population. That is also the reason why two sites (118 K and 162 L) at the *ND6* gene were identified with more than 96% probability under positive selection by nested model comparison, while likelihood ratio tests of M1 vs M2 and M7 vs M8 were not significant ( $2\Delta\ln L_{M1 \text{ vs } M2} = 4.2336$  with  $0.1 < P < 0.5$ ;  $2\Delta\ln L_{M7 \text{ vs } M8} = 5.1531$  with  $0.05 < P < 0.1$ ) (Table 2).

### CONCLUSIONS

The lowest genetic diversity and significant negative Tajima's  $D$  in the high-altitude population, the non-directional selection signal, the high proportion of plateau horse samples on the selected branches, and the two potential selected sites, all lead us to suggest that the *ND6* gene has undergone adaptive evolution to the high-altitude environment in most Chinese plateau horses. In order to test whether adaptive evolution is unique to the horse mitochondrial *ND6* gene, the same analyses were performed on the mitochondrial genomes (T. Ning, H. Xiao, J. Li, S. Hua, et al., unpublished data); however, we did not find a similar pattern in any other genes. This research helps demonstrate that severe environments can compel adaptive evolution in mitochondrial DNA.

### ACKNOWLEDGMENTS

We thank Dr. S.Y. Chen, Y.W. Miao and J.J. Gao for helpful comments on the manuscript, and we are also grateful to G.D. Wang, Z.H. Zhu, W.P. Zhou, H.B. Xie, and S.F. Wu for technical assistance. Research supported by the National Basic Research Program of China (973 program: #2007CB815700), the National Natural Science Foundation of China (#30760038), the Chinese Academy of Sciences (#KSCX2-YW-N-018), and the Bureau of Science and Technology of Yunnan Province.

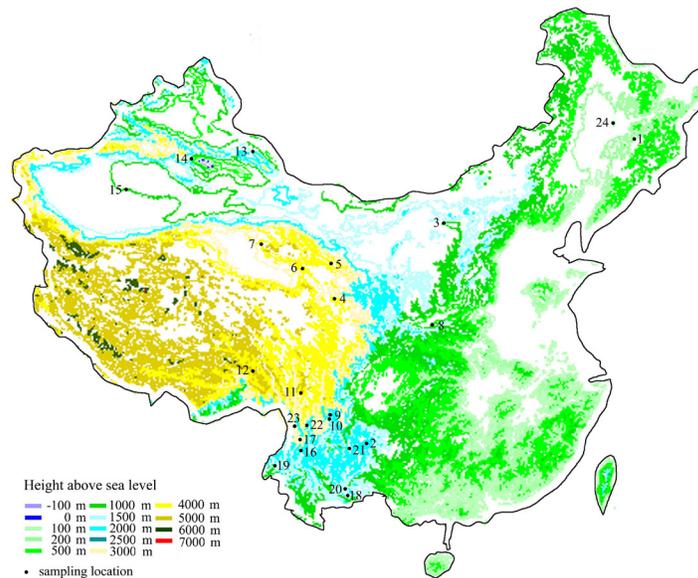
**REFERENCES**

- Beall CM (2003). High-altitude adaptations. *Lancet* 362 (Suppl): s14-s15.
- Creevey CJ and McInerney JO (2002). An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. *Gene* 300: 43-51.
- Excoffier L, Laval G and Schneider S (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform (Online)* 1: 47-50.
- Gering EJ, Opazo JC and Storz JF (2009). Molecular evolution of cytochrome b in high- and low-altitude deer mice (genus *Peromyscus*). *Heredity* 102: 226-235.
- Hassanin A, Ropiquet A, Couloux A and Cruaud C (2009). 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.* 68: 293-310.
- Hudson RR, Slatkin M and Maddison WP (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583-589.
- Lopez-Barneo J, Pardal R and Ortega-Saenz P (2001). Cellular mechanism of oxygen sensing. *Annu. Rev. Physiol.* 63: 259-287.
- Monge C and Leon-Velarde F (1991). Physiological adaptation to high altitude: oxygen transport in mammals and birds. *Physiol. Rev.* 71: 1135-1172.
- Nielsen R (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39: 197-218.
- Outram AK, Stear NA, Bendrey R, Olsen S, et al. (2009). The earliest horse harnessing and milking. *Science* 323: 1332-1335.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
- Schmidt-Nielsen K (1997). Effect of High Altitude. In: *Animal Physiology: Adaptation and Environment*. Cambridge University Press, Cambridge, 209-210.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.
- Weber RE (2007). High-altitude adaptations in vertebrate hemoglobins. *Respir. Physiol. Neurobiol.* 158: 132-142.
- Wickler S and Greene H (2003). The horse and high altitude. *Clin. Tech. Equine Pract.* 2: 231-237.
- Xu S, Luosang J, Hua S, He J, et al. (2007). High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. *J. Genet. Genomics* 34: 720-729.
- Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13: 555-556.

**Table 1 (Supplementary).** Sampling information for the 509 Chinese domestic horses.

Group	Breed	Sampling location	Mean altitude	Population	Sample size
1	Jilin	Song-Liao Plain, Jilin	110-200 m	LP	17
2	Weining	Weining, Guizhou	2200 m	HP	15
3	Inner Mongolian	Inner Mongolian	1200 m	MP	52
4	Tongde	Tongde, Qinghai	3700 m	HP	31
5	Delingha	Delingha, Qinghai	2980 m	HP	15
6	Wulan	Wulan, Qinghai	2960 m	HP	17
7	Qilian	Qilian, Qinghai	3170 m	HP	11
8	Guanzhong	Guanzhong Plain, Shaanxi	500 m	LP	33
9	Luhe	Xichang, Sichuan	1500 m	MP	4
10	Liangshan	Xichang, Sichuan	1500 m	MP	12
11	Tibetan	Zoige Plateau, Sichuan	3500 m	HP	9
12	Tibetan	Xizang	4200 m	HP	7
13	Balikun	Balikun county, Xinjiang	1650 m	MP	44
14	Hejing	Haermodun town, Xingjiang	1200 m	MP	21
15	Hetian	Hetian, Xingjiang	1300-1450 m	MP	7
16	Dali	Heqing county, Yunnan	2980 m	HP	14
17	Lijiang	Yulong county, Yunnan	3100 m	HP	31
18	Maguan	Muchang town, Yunnan	1200 m	MP	53
19	Tengchong	Tengchong county, Yunnan	1640 m	MP	31
20	Wenshan	Wenshan county, Yunnan	1600 m	MP	16
21	Wumeng	Wumeng mountain, Yunnan	3200 m	HP	28
22	Yongning Tibetan	Ninglang county, Yunnan	2800 m	HP	5
23	Zhongdian Tibetan	Diqing, Yunnan	3300 m	HP	15
24	Shuangcheng	Shuangcheng, Heilongjiang	165 m	LP	21

HP = altitude >2200 m; MP = altitude 1200-1700 m; LP = altitude <900 m.



**Figure 1 (Supplementary).** Sampling locations and altitude situation of Chinese domestic horses. Numbers by the dark dots are the same group numbers as in Supplementary Table 1.