

# Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*

B. Wallner\*, G. Brem<sup>†</sup>, M. Müller\* and R. Achmann<sup>†</sup>

\*Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, Veterinärplatz, Wien, Austria.

<sup>†</sup>Ludwig-Boltzmann-Institut für immuno-, zyto- und molekulargenetische Forschung, Veterinärmedizinische Universität Wien, Veterinärplatz, Wien, Austria

## Summary

The phylogenetic relationship between *Equus przewalskii* and *E. caballus* is often a matter of debate. Although these taxa have different chromosome numbers, they do not form monophyletic clades in a phylogenetic tree based on mtDNA sequences. Here we report sequence variation from five newly identified Y chromosome regions of the horse. Two fixed nucleotide differences on the Y chromosome clearly display Przewalski's horse and domestic horse as sister taxa. At both positions the Przewalski's horse haplotype shows the ancestral state, in common with the members of the zebra/ass lineage. We discuss the factors that may have led to the differences in mtDNA and Y-chromosomal observations.

**Keywords** *Equus caballus*, *Equus przewalskii*, phylogeny, Y chromosome.

Investigation of genetic variation has been widely used to clarify the evolution of the genus *Equus*. The deepest node in the phylogeny of the equine species is a divergence between the lineage leading to true horses such as domestic horse (*Equus caballus*) and Przewalski's wild horse (*E. przewalskii*) and the ancestral lineage leading to zebras and asses (Oakenfull *et al.* 2000). Many studies have specifically focused on the phylogenetic relationship between the Przewalski's horse and the domestic horse but the findings are controversial. Przewalski's horses have two more chromosomes than domestic horses, which is suggested as being an ancestral feature (Benirschke *et al.* 1965). Morphological characteristics (Groves 1994) and genetic studies of blood markers (Bowling & Ryder 1987) also distinguish these taxa, but there is substantial overlap in the mtDNA sequence variation of *E. caballus* and *E. przewalskii* (Ishida *et al.* 1995; Oakenfull & Ryder 1998; Oakenfull *et al.* 2000; Vila *et al.* 2001). Based on mitochondrial DNA (mtDNA) phylogeny, the hypothesis that Przewalski's horse is the ancestral type could not be supported.

Przewalski's horse became extinct in the wild in the 1960s and the captive population descends from 13 individuals. Of these 13 animals one was a domestic horse and one a domestic/Przewalski hybrid. According to official

studbook records, four maternal lineages survived, none of which has descended from domestic founders. Sequencing showed that only two mtDNA types exist in these four lineages (Oakenfull & Ryder 1998). Oakenfull *et al.* (2000) claim that the mtDNA similarities between domestic and Przewalski's could be as a result of ancestral polymorphisms before the wild forms diverged or because of gene flow between the two types. Interbreeding is known to have occurred before the 1960s, whereas predominant mating between Przewalski's stallions and domestic mares has been observed (Volf 1994).

In this study we investigate the genetic relationship between *E. caballus* and *E. przewalskii* paternal lines, based on Y-chromosomal sequence variation. The Y chromosome, except for its pseudoautosomal region, is transmitted paternally as a linkage group, which permits the possibility of contrasting male-specific histories of populations with that of females, revealed by mtDNA.

We applied representational difference analysis (RDA) using the *Bgl*II restriction enzyme and oligonucleotides, as described in Lisitsyn *et al.* 1993, to isolate Y-specific DNA sequences from the horse. Genomic DNA from a male and a female Lipizzan horse were used as tester and driver, respectively. Selective polymerase chain reaction (PCR) amplification and hybridization was performed in three cycles and subcloned PCR products were analysed by sequencing on an ABI 377 sequencer. We assigned five randomly selected non-coding sequences (Eca-Y3B1, Eca-Y3B8, Eca-Y3B12, Eca-Y2B17, Eca-Y3B19, GenBank accession numbers G72335–G72339) covering a total length of 2163 nucleotides to the horse Y chromosome based on male-specific PCR test.

## Address for correspondence

B Wallner, Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, Veterinärplatz 1, A-1210 Wien, Austria.  
E-mail: wallner@i122server.vu-wien.ac.at  
Present address: R. Achmann, genteQ GmbH, Falkenried 88, D-20251 Hamburg, Germany.

Accepted for publication 1 July 2003

**Table 1** Observed number of substitutions in five Y chromosome regions (1886 bp) above the diagonal, average pairwise distance using maximum likelihood in percentage with the HKY model of sequence evolution below the diagonal.

Sequence divergence among equine Y chromosomes									
	Domestic horse	Przewalski's horse	Hartmann's zebra	Damara zebra	Grant's zebra	Grevyi's zebra	Kiang	Onager	Donkey
Domestic horse	2	29	29	28	30	31	32	32	
Przewalski's horse	0.107	27	27	26	28	29	30	30	
Hartmann's zebra	1.569	1.459	4	5	9	10	11	11	
Damara zebra	1.568	1.459	0.212	1	9	10	11	11	
Grant's zebra	1.514	1.404	0.266	0.053	8	9	10	10	
Grevyi's zebra	1.624	1.514	0.479	0.479	0.425	11	12	12	
Kiang	1.679	1.569	0.532	0.532	0.479	0.586	5	7	
Onager	1.735	1.625	0.586	0.586	0.533	0.640	0.266	8	
Donkey	1.734	1.624	0.586	0.586	0.532	0.640	0.372	0.426	

\*Indicates breeds investigated for sequence regions Eca-Y2B17 and Eca-Y3B19 only.

Domestic horses breeds: Akhal-Teke, Andalusian, Appaloosa\*, Arabian, Austrian Warmblood, Barb\*, Connemara\*, Icelandic Horse, Kladruber, Lipizzaner ( $n = 8$ ), Mangalarga Marchador\*, Miniature\*, Mongolian native horse, New Forest Pony\*, Noric, Norwegian Fjord\*, Oldenburger\*, Old Wuerttemberger\*, Paint\*, Pinto\*, Quarter Horse, Saddlebred\*, Shetland Pony, Shagya Arabian\*, Shire\*, Tarpan-like horse, Thoroughbred, Tinker\*, Trakehner, Trotter\*.

We examined Y chromosome sequence variation in five regions over a total length of 1887 bp by direct sequencing of gel purified PCR products. Blood samples from 21 purebred domestic horse stallions (*E. caballus*) representing 14 different breeds (Table 1) were investigated. The geographical origins of the samples correspond mostly to the main breeding area.

Additionally we analysed Przewalski's horses (*E. przewalskii*,  $n = 10$ ), donkeys (*E. asinus*,  $n = 2$ ) and one member each of the taxon Damara zebra (*E. quagga antiquorum*), Grant's zebra (*E. quagga boehmi*), Hartmann's mountain zebra (*E. zebra hartmannae*), Grevyi's zebra (*E. grevyi*), Onager (*E. hemionus onager*) and Kiang (*E. kiang*). The PCR was performed in a 50  $\mu$ l volume containing 0.5  $\mu$ M of each primers (Eca-Y2B17: 5'-TTCAGTCCTGCTTTCTCCTCA and 5'-CAGGATGTGCCATGTGATTG; Eca-Y3B1: 5'-TGGGTTAATGGGATTTGGTG and 5'-CAAGCACAGCTCTGTATCAA; Eca-Y3B8: 5'-CCCAAGTTCCTTGCCATC and 5'-AAATTGAAAGAGCCCCAAAG; Eca-Y3B12: 5'-GGGAGGCACTGGAAAGTACA and 5'-GGTGGAGGAATCAGCTGGAG; Eca-Y3B19: 5'-AAGCCTTTCATGGAAATTGG and 5'-TTACGCAGACATCCTGGACA), 1 $\times$  AmpliTaq buffer, 2 mM MgCl<sub>2</sub>, 0.25 mM each dNTP's, 2 U AmpliTaq Gold and 40 ng genomic DNA. The DNA was initially denatured at 95 °C for 10 min followed by 35 cycles of 30 s 95 °C, 40 s 58 °C and 90 s 72 °C.

We found no polymorphic nucleotide sites on the total length of 1887 nucleotides during the initial survey of domestic horse stallions. Simultaneously, we observed no intraspecific variation within the Przewalski's horse and the donkey. This observation is consistent with the general finding of a low level of intraspecific sequence variability on the mammalian Y chromosome (Shen et al. 2000). The maximum likelihood distance between taxa, calculated with Tree-puzzle 5.0 (<http://www.tree-puzzle.de>, see Table 1),

ranged from 0.053% (Damara zebra – Grant's zebra) to 1.735% (domestic horse – Onager). The domestic horse differed from Przewalski's horse by one base substitution in each of two regions (Eca-Y2B17 and Eca-Y3B19). At both polymorphic sites the Przewalski's horse showed the ancestral state in common with the taxa of the ass/zebra lineage (Fig. 1). To confirm the fixed nucleotide differences between the domestic and Przewalski's horse, we sequenced one stallion from 16 additional domestic horse breeds for region Eca-Y2B17 and Eca-Y3B19 (Table 1). All 16 samples were found to be identical in sequence and showed the apomorphic state of the domestic horse at both nucleotide sites.

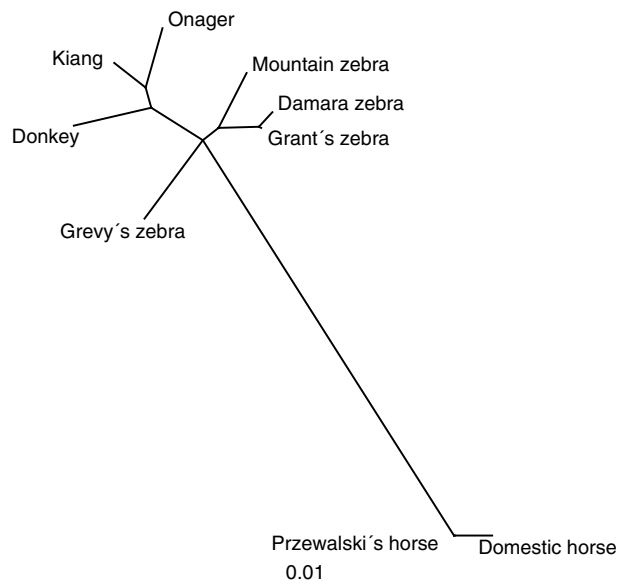
Patterns of nucleotide sequence diversity of the horse Y chromosome suggest that the Przewalski's haplotype is distinct to that of the domestic horse. This distinction is also evident in the maximum likelihood tree of Y chromosome sequences where Przewalski's horse and domestic horse are shown as sister taxa (Fig. 2), but form a distinct cluster to the other equids. However, the haplotype of *E. przewalskii* is more similar to that of their common ancestor.

Under the assumption of the molecular clock we estimated the rate of nucleotide substitution on the Y chromosome based on the equation  $r = 2/dt$ , where  $d$  is the genetic divergence between species and  $t$  is their divergence time. Assuming the horse and the donkey separated 2 to 3.9 million years ago (Vila et al. 2001), we calculated a substitution rate of  $4.34 \times 10^{-9}$  to  $2.22 \times 10^{-9}$  substitutions per site per year on the Y chromosome. The time for the split of the domestic and Przewalski's horse is therefore between 123 300 and 241 100 years ago (95% confidence limits for the Poisson parameter  $\lambda = 2$ : 21 700–408 400 years ago and 42 400–798 400 years ago, respectively (Zar 1999). As our estimate for the split between *E. przewalskii* and *E. caballus* predates by over 100 000 years the proposed domestication event 6000 years ago (Clutton-Brock 1999),

	LocusYB1												LocusYB8																								
	7	7	9	9	5	8	9	0	3	3	8	0	3	3	6	7	5	7	7	7	7	7	7	8	8	8	8	8	8	6	9	0	8	6	4		
Domestic horse	4	8	3	7	4	0	0	5	0	5	9	7	3	8	3	5	1	3	4	5	6	7	8	9	0	1	2	3	4	9	2	9	3	4	7		
Przewalski's horse	A	T	C	T	G	C	A	G	C	G	C	C	C	C	C	C	C	G	-	-	-	-	-	-	-	-	-	-	-	A	A	T	C	A	A		
Hartmann's zebra	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Damara zebra	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Grant's zebra	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Grevy's zebra	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Kiang	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Onager	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Donkey	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

**Figure 1** Polymorphic sites between five Y-chromosomal regions of several equine species (1887 in total). Invariant sites are marked with dots and dashes refer indels. Diagnostic sites separating the domestic and Przewalski's horses are shown in bold. All sequences are deposited on Genbank (Accession Numbers AY225598, AY228242–AY228280).

	LocusYB12												LocusYB17												LocusYB19							
	3	4	7	4	8	6	8	8	9	0	0	7	7	5	5	5	6	8	1	2	4	7	2	2	5	9	4	5	6	9	1	
Domestic horse	3	9	3	9	3	3	1	3	5	6	9	0	0	6	7	8	9	0	8	3	8	0	4	9	7	1	0	5	4	0	7	4
Przewalski's horse	C	G	C	C	C	C	T	C	C	C	C	T	A	-	-	-	-	-	-	G	T	C	C	C	C	G	A	C	G	C	C	
Hartmann's zebra	T	A	T	A	.	.	.	T	C	.	.	.	.	G	T	G	G	T	G	C	T	T	.	.	.	.	C	T	A	.	.	
Damara zebra	.	A	T	A	.	.	.	T	C	.	.	C	.	G	T	G	G	T	G	.	C	T	T	.	.	.	C	T	A	.	.	
Grant's zebra	.	A	T	A	.	.	.	T	C	.	.	C	.	G	T	G	G	T	G	.	C	T	T	.	.	.	C	T	A	.	.	
Grevy's zebra	.	A	T	A	.	.	.	T	C	T	.	.	.	C	G	T	G	G	T	G	.	C	T	T	.	.	C	T	A	.	.	
Kiang	.	A	T	A	.	.	.	T	C	.	.	.	.	G	T	G	G	T	G	.	C	T	T	.	.	.	C	T	A	.	.	
Onager	.	A	T	A	.	.	.	T	C	.	.	.	.	G	T	G	G	T	G	.	C	T	T	.	.	.	C	T	A	.	.	
Donkey	.	A	T	A	.	.	.	T	C	.	.	.	.	G	T	G	G	T	G	.	C	T	T	.	.	.	C	T	A	.	.	



**Figure 2** Unrooted maximum likelihood tree of Y chromosome sequences (1887 bp) from nine equid species. The root of the donkey–zebra group is a trichotomy. The rest of the grouping is supported by 100% puzzle support (based on 50 000 replications).

our data suggest that the extant Przewalski's horse population does not represent the wild stock from which the modern horse was bred (Bowling & Ruvinsky 2000).

No diagnostic differences have been found in former studies between Przewalski's and domestic horses for mtDNA. In contrast we detected fixed differences on the Y chromosome. Therefore, based on our sample, we observed no evidence for male mediated admixture between Przewalski's and domestic horses.

The differences detected between mtDNA and Y markers may be explained by female mediated gene flow of domestic horse lineages into the Przewalski's population. Nevertheless, differential effective population sizes in females and males could also account for this pattern. The ratio of

female to male founders is skewed in favour of females in polygynous species, such as the wild horses. Intensive breeding strategies have further dramatically reduced male effective population size in the domestic horse (Nowak 1991). In addition to the severe bottleneck that occurred in the Przewalski's population, this may result in a total depletion of genetic diversity and, due to drift, leading to fixation of different Y-chromosomal haplotypes.

## Acknowledgements

We thank W. Zimmermann, C. Waltzer, H. Korb, E. Dvorak, the Zoological Garden Munich and L. Chemnick from the Zoological Society of San Diego for samples. We are grateful to M. Wigler for a detailed RDA protocol. Thanks to G. Muir and C. Vogl for comments on the manuscript. This study was financially supported by the Hochschuljubiläumsstiftung der Stadt Wien and Agrobiogen GmbH.

## References

- Benirschke K., Malouf N., Low R. & Heck H. (1965) Chromosome complement: differences between *Equus caballus* and *Equus przewalskii*, Poliakoff. *Science* **148**, 382–3.
- Bowling A.T. & Ruvinsky A. (2000) Genetic aspects of domestication, breeds and their origins. In: *The Genetics of the Horse* (Ed. by Bowling, A.T. & Ruvinsky, A.) pp. 25–51. CABI Publishing, Wallingford.
- Bowling A. & Ryder O.A. (1987) Genetic studies of blood markers in Przewalski's horses. *Journal of Heredity* **78**, 75–80.
- Clutton-Brock J.A. (1999) *A natural History of Domesticated Mammals*. Cambridge University Press, Cambridge.
- Groves C. (1994) Morphology, habitat and taxonomy. In: *Przewalski's horse. The history of an endangered species*. (Ed. by Boyd L. & Houpt K.), pp. 39–59. State University Press, New York.
- 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 Molecular Evolution* **41**, 180–8.

- Lisitsyn N., Lisitsyn N. & Wigler M. (1993) Cloning the differences between two complex genomes. *Science* **259**, 946–51.
- Nowak R.M. (1991) *Walker's Mammals of the World*. Johns Hopkins University Press, Baltimore.
- Oakenfull E.A. & Ryder, O.A. (1998) Mitochondrial control region and 12SrRNA variation in Przewalski's horse (*Equus przewalskii*). *Animal Genetics* **29**, 456–9.
- Oakenfull A., Lim A. & Ryder O. (2000) A survey of equid mitochondrial DNA: implications for the evolution, genetic diversity and conservation of *Equus*. *Conservation Genetics* **1**, 341–55.
- Shen P., Wang, F., Underhill P.A. *et al.* (2000) Population genetic implications from sequence variation in four Y chromosome genes. *Proceedings of the National Academy of Sciences of the USA* **97**, 7354–9.
- Vila C., Leonard J., Götherström A., Marklund S., Sandberg K., Liden K., Wayne R. & Ellegren H. (2001) Widespread origins of domestic horse lineages. *Science* **291**, 474–7.
- Volf J. (1994) *Das Urwildpferd*, 4th edn. Spektrum Akademischer Verlag, Heidelberg.
- Zar J.H. (1999) *Biostatistical analysis*. Prentice-Hall, New Jersey.