J Mol Evol (1995) 41:180–188

JOURNAL OF MOLECULAR EVOLUTION

Mitochondrial DNA Sequences of Various Species of the Genus *Equus* with Special Reference to the Phylogenetic Relationship Between Przewalskii's Wild Horse and Domestic Horse

Nobushige Ishida,[1] Tsendsuren Oyunsuren,[2] Suguru Mashima,[3] Harutaka Mukoyama,[1] Naruya Saitou[4]

[1] Laboratory of Molecular and Cellular Biology, Equine Research Institute, Japan Racing Association, 27-7, Tsurumaki 5-chome, Setagaya-ku, Tokyo 154, Japan
[2] Department of Molecular Biology, Institute of Biotechnology, Enkhtaivan ave 54, Ulaanbaatar 210351, Mongolia
[3] Section of Molecular Genetics, Laboratory of Racing Chemistry, 4-37-6, Kamiyoga, Setagaya-ku, Tokyo 158, Japan
[4] Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima 411, Japan

Received: 7 November 1994 / Accepted: 27 February 1995

Abstract. The noncoding region between $tRNA^{Pro}$ and the large conserved sequence block is the most variable region in the mammalian mitochondrial DNA D-loop region. This variable region (ca. 270 bp) of four species of *Equus,* including Mongolian and Japanese native domestic horses as well as Przewalskii's (or Mongolian) wild horse, were sequenced. These data were compared with our recently published Thoroughbred horse mitochondrial DNA sequences. The evolutionary rate of this region among the four species of *Equus* was estimated to be $2–4 \times 10^{-8}$ per site per year. Phylogenetic trees of *Equus* species demonstrate that Przewalskii's wild horse is within the genetic variation among the domestic horse. This suggests that the chromosome number change (probably increase) of the Przewalskii's wild horse occurred rather recently.

Key words: Mitochondrial DNA — Noncoding D-loop region — Phylogenetic tree — Przewalskii's wild horse — *Equus caballus* — Evolutionary rate

Introduction

Tracing the evolution of the horse family *Equidae* from eohippus (genus *Hyracotherium*) to the modern genus *Equus* is a classic achievement of paleontology (e.g., Simpson 1951; Groves and Willoughby 1981; MacFadden 1992). Most members of the horse family are now extinct, and all the living Equidae species represent a single genus, *Equus*. The phylogenetic relationship within the species of this genus is, however, not well understood. According to recent literature, *Equus* is classified into four subgenera with seven species: subgenus *Equus* includes *E. caballus* (domestic horses and Przewalskii's or Mongolian wild horse); subgenus *Asinus* includes *E. asinus* (donkeys), *E. hemionus* (onagers), and *E. kiang* (kiang); subgenus *Dolichohippus* has only *E. grevyi* (Grevy's zebra); and *E. zebra* (mountain zebra), *E. burchelli* (Burchell's or plains zebra), and the recently extinct *E. quagga* (quagga) are classified into subgenus *Hippotigris* (Nowak and Paradiso 1983).

Because its chromosome number (2n = 66) is different from that (2n = 64) of domestic horses (Benirschke et al. 1965), some researchers classify Przewalskii's wild horse as a distinct species, *E. przewalskii* (e.g., Ryder et al. 1978). It is generally considered that two pairs of acrocentric chromosomes in the ancestral horse population (2n = 66; Przewalskii's wild horse type) changed to one pair of submetacentric chromosomes by Robertsonian fusion, and this produced the modern horse chromosome number (2n = 64) (Benirschke et al. 1965; Ryder et al. 1978). By comparison of equine hemoglobin amino acid sequences (Clegg 1974) and serum protein studies (Ryder et al. 1979; Kaminski 1979), a close relationship between domestic horses and Przewalskii's

Correspondence to: N. Ishida

wild horses has been shown. Genetic material of these two species is so similar that their hybrids are fertile (Ryder et al. 1978; Trommerhausen-Smith et al. 1979). Therefore, there is uncertainty on the species status of *E. przewalskii,* and it is tentatively included into *E. caballus* in the following.

Mammalian mitochondrial DNA (mtDNA) has a rate of nucleotide substitution five to ten times as high as nuclear DNA (Brown et al. 1979). George and Ryder (1986) showed, on the basis of their several equine mitochondrial DNA restriction-site maps, that the sequence divergence between Przewalskii's wild horse and domestic horse was within the intraspecific variation. George and Ryder (1986) also examined five other *Equus* species, but the phylogenetic relationship among those *Equus* species was not clarified.

Several studies have suggested that the most rapidly evolving part of the mitochondrial genome is the noncoding control region which contains the D-loop (Walberg and Clayton 1981; Chang and Clayton 1985). The substitution rate in the human D-loop region has been estimated to be 2.8–5 times higher than the rate for the remainder of the mitochondrial genome (Aquadro and Greenberg 1982; Cann et al. 1984). Recently we sequenced the whole D-loop region (about 1,100 bp) for three Thoroughbred horse mtDNAs (Ishida et al. 1994) and found the noncoding region between $tRNA^{Pro}$ and the large conserved sequence block was most variable among the three Thoroughbred individuals. In the present study, we sequenced the corresponding six equine mtDNA D-loop regions and conducted a phylogenetic analysis of the genus *Equus*. Special reference will be made to the phylogenetic relationship between Przewalskii's (or Mongolian) wild horse and domestic horse.

Materials and Methods

Isolation of Genomic DNA. The blood of *E. zebra* and *E. asinus* was obtained through the courtesy of Yumemigasaki Zoological Park, Kawasaki. The blood of *E. grevyi* was a gift from Tama Zoological Park, Tokyo. The blood of a Przewalskii's wild horse (International Studbook Number 880) and a Japanese native horse were from the Equestrian Park of Japan Racing Association, Tokyo. The blood sample of a Mongolian native horse was collected in Mongolia. Genomic DNA was isolated from peripheral blood of those six individuals by using the standard techniques (Sambrook et al. 1989). The chromosome numbers (2n) of those sampled individuals were 64, 66, 62, 46, and 32 for Mongolian and Japanese native horses, Przewalskii's wild horse, *E. asinus, E. grevyi,* and *E. zebra,* respectively (K. Hirota personal communication, 1993), compatible with previous studies (e.g., Ryder et al. 1978).

Sequencing. PCR primers were designed to amplify the region between $tRNA^{Pro}$ and the large conserved sequence block on the basis of three Thoroughbreds' DNA sequences (Ishida et al. 1994). Primer sequences are 5′-gctgaaattctacttaaa-3′ and 5′(biotin labeled)-agttggaaggttgctgatt-3′. PCR was performed using 600-ng genomic DNA, 1.2 μM of each primer in 75 μl 1 × PCR buffer containing 1.25 mM of each dNTP, and 2.5U Amplitaq Polymerase (Perkin Elmer Cetus). Amplification conditions were 96°C for 2 min, followed by 35 cycles of 94°C 1 min, 58°C 1 min, and 72°C 1 min, and final 10-min extension at 72°C. PCR product was purified with MicroSpin Columns (Pharmacia). The alkali-denatured single-stranded DNA was collected with Dynabeads (Dynal). Sequence analysis was done by the dideoxyribonucleotide chain termination method (Sanger et al. 1977), using the Automated Laser Fluorescent (ALF) sequencer (Pharmacia).

Data Analysis. Raw sequence data were analyzed by using software DNASIS-Mac (Hitachi Software Engineering), and CLUSTAL V (Higgins et al. 1992) was used for multiple sequence alignment. Phylogenetic trees were constructed by using the neighbor-joining method (Saitou and Nei 1987) and the maximum parsimony method (Fitch 1977). Computer packages MEGA (Kumar et al. 1993) and PAUP (Swofford 1993) were used for assisting the neighbor-joining and the maximum parsimony analysis, respectively. Numbers of nucleotide substitutions between sequences were estimated by using Tamura and Nei's (1993) substitution model with a gamma distribution ($\alpha = 0.5$), because that model seems to be suitable for estimating the hypervariable region of mitochondrial DNA D-loop (Kumar et al. 1993).

Results and Discussion

Nucleotide Differences

Approximately 270 bp of noncoding regions of mtDNA for *E. zebra, E. grevyi, E. asinus,* and three *E. caballus* (Mongolian native horse, Japanese native horse, and Przewalskii's wild horse) were sequenced using the PCR direct sequencing method. Figure 1 presents the multiple-aligned sequences of this region for all studied individuals as well as the three Thoroughbreds of Ishida et al. (1994). Five nucleotide positions (indicated by an asterisk at the bottom of each sequence block) were assumed to contain at least one gap out of 274 positions. Nucleotide differences were more frequently observed near the 5′ region (close to $tRNA^{Pro}$) of sequenced DNA than near the 3′ region, and a total of 60 nucleotide sites were variable among the 269 positions without any gap. The pairwise sequence differences (transitional and transversional differences among those 269 positions) are shown in Table 1. It is clear that transitions outnumbered transversions.

When we restrict the comparison into the six *E. caballus* sequences (three Thoroughbreds, Mongolian native horse, Japanese native horse, and Przewalskii's wild horse), only Przewalskii's wild horse has one insertion at position 103, and 271 nucleotide positions do not have any gap. Sequence data for these 271 positions were used for the phylogenetic analysis of *E. caballus* sequences. Nineteen sites were variable among those 271 positions, and a majority of the nucleotide differences were transitional ones.

Phylogenetic Tree of the Six E. caballus *Sequences*

Figure 2 is a phylogenetic tree of six *E. caballus* sequences. After the neighbor-joining tree was constructed

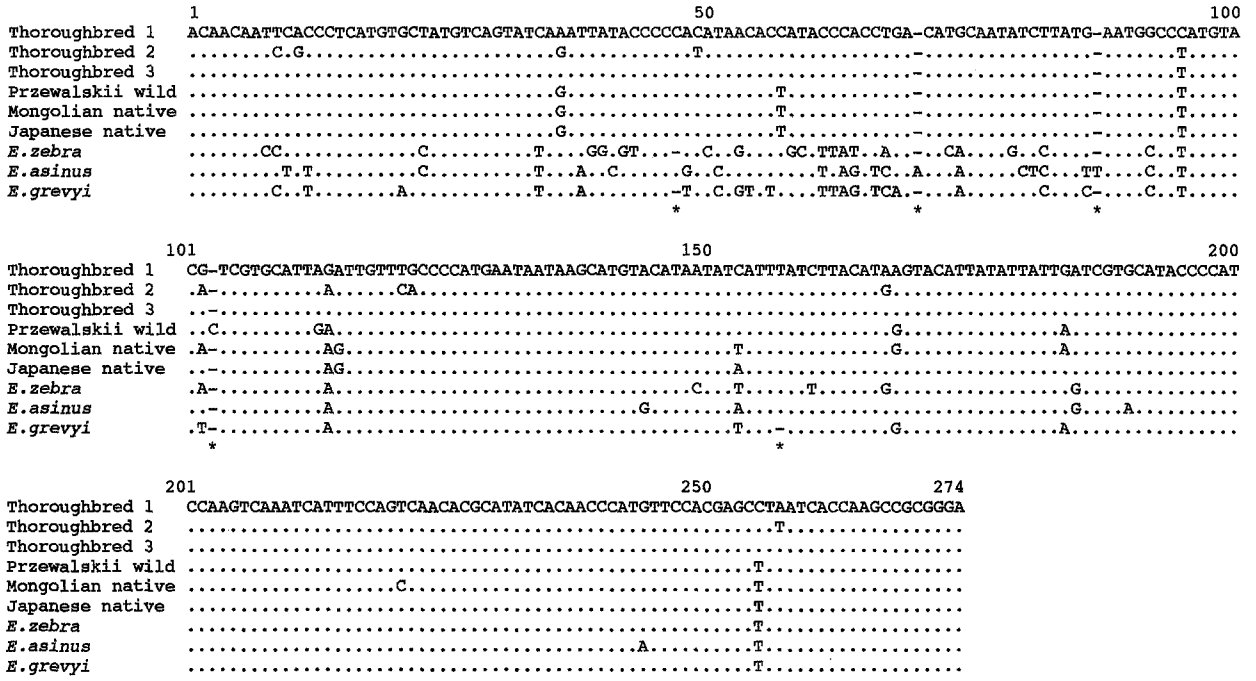

Fig. 1. An alignment for nine equine mitochondrial DNA sequences. Three already-known Thoroughbred sequences (Thoroughbred 1, 2, and 3) of Ishida et al. (1994) were included with the six sequences determined by the present study; Przewalskii's wild horse, Mongolian native horse, Japanese native horse, *E. zebra, E. asinus,* and *E. grevyi.* Sequence identity is indicated by a *dot,* and gap is indicated by a *hyphen.* Five nucleotide positions that include gaps are indicated with *asterisks* at the bottom.

Table 1. Nucleotide differences among the nine equine mtDNA sequences[a]

Sequence name (abbr.)	T1	T2	T3	Pr	Mo	Jp	Zb	As	Gv
Thoroughbred 1 (T1)	—	1	0	0	0	1	7	5	6
Thoroughbred 2 (T2)	10	—	1	1	1	2	8	6	7
Thoroughbred 3 (T3)	1	9	—	0	0	1	7	5	6
Przewalskii wild (Pr)	8	12	7	—	0	1	7	5	6
Mongolian native (Mo)	11	13	10	5	—	1	7	5	6
Japanese native (Jp)	6	10	5	4	4	—	8	4	7
E. zebra (Zb)	24	24	23	26	25	23	—	8	9
E. asinus (As)	22	28	21	24	26	22	25	—	3
E. grevyi (Gv)	22	25	20	20	20	21	21	19	—

[a] Numbers below and above diagonal are transitional and transversional differences, respectively, among the 269 nucleotide positions without gaps (see Fig. 1)

using the original pairwise distances, branch lengths (number of nucleotide substitutions per site) were multiplied the number of nucleotides was compared (271 in the present case), and integer values were obtained after rounding. These integer-value branch lengths were compared with those estimated by using the maximum parsimony principle applied to the neighbor-joining tree, and the final branch lengths were determined. This procedure follows that of Nerurker et al. (1993). A total of 24 nucleotide substitutions were estimated in the tree of Fig. 2 by applying this procedure. The location of the root was assumed to be equal to that of Fig. 3, in which three other *Equus* species sequences were also used.

Numbers above the three internal branches are bootstrap probabilities (in %) based on 1,000 bootstrapped neighbor-joining trees. Relatively high bootstrap probabilities of 90% and 88% were observed for the clustering of Thoroughbred 1 and 3 sequences and that of three Thoroughbred sequences, respectively, while the clustering of Przewalskii's wild horse and Mongolian native horse showed a 61% probability.

We found a unique most parsimonious tree by using the branch-and-bound option of PAUP, and the topology of that tree was identical with that of the neighbor-joining tree (Fig. 2). The maximum parsimony tree required 22 nucleotide substitutions, two substitutions less than that for the neighbor-joining tree. Numbers in parentheses below the three internal branches of the tree of

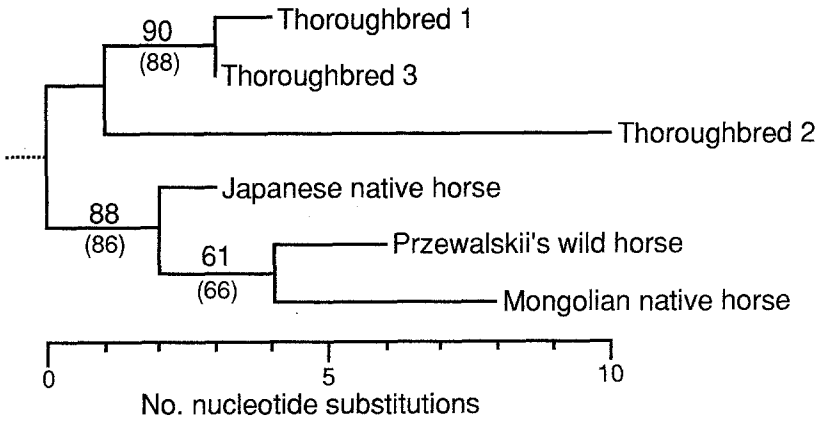

Fig. 2. A phylogenetic tree of the six *Equus caballus* sequences constructed by using the neighbor-joining method. *Figures above and below (in parentheses)* the three internal branches are bootstrap probabilities (in %) based on 1,000 bootstrapped trees when the neighbor-joining method and the maximum parsimony method were used, respectively.

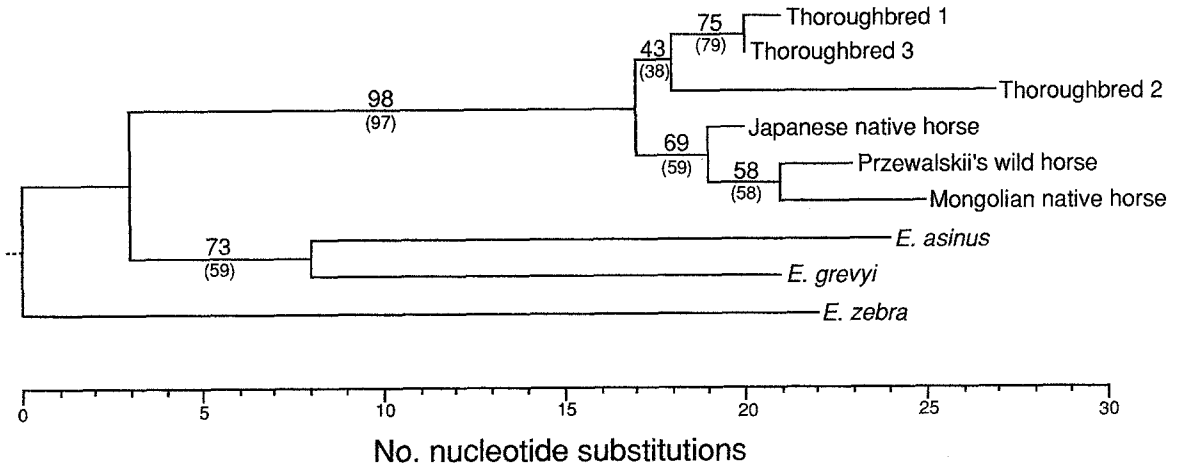

Fig. 3. A phylogenetic tree of nine *Equus* sequences constructed by using the neighbor-joining method. See text for the rooting procedure. *Figures above and below (in parentheses)* the three internal branches are bootstrap probabilities (in %) based on 1,000 bootstrapped trees when the neighbor-joining method and the maximum parsimony method were used, respectively.

Fig. 2 are bootstrap probabilities (in %) based on 1,000 bootstrapped maximum parsimony trees. The bootstrap values are more or less the same as those for the neighbor-joining tree.

Table 2 shows a list of estimated change in nucleotides mapped to the tree of Fig. 2. Two substitutions out of 24 were nonparsimonious estimation. Three nucleotide substitutions are assumed at position 153, and all of them are transversions, while the most parsimonious explanation for the nucleotide configuration of this site under the tree of Fig. 2 requires two substitutions involving one transversion (shown in parentheses in Table 2). Another nonparsimonious substitution is assumed at the branch going to the Thoroughbred 2 sequence. (Its position is designated by "?" in Table 2.) This extra substitution was estimated apparently because of the use of Tamura and Nei's (1993) substitution model with a gamma distribution. If this is a real one, it was probably a transition. Under this assumption, four out of 24 substitutions are transversions, or if we choose the parsimonious explanation at position 153, two (C → A at position 153 and A → T at position 257) out of 23 are transversions. Depending on either case, transitions occurred 5–10.5 times more than transversions. This transition/transversion ratio is slightly lower than that (ca. 15) observed in the human mtDNA D-loop region (Vigilant et al. 1991; Tamura and Nei 1993). Among the 19 (or 20) transitions, 7 (or 8) are changes between pyrimidines (T ⟷ C) and the remaining 12 are those between purines (A ⟷ G).

Phylogenetic Tree of the Nine Equus *Sequences*

Figure 3 is a phylogenetic tree of nine *Equus* sequences constructed by using the neighbor-joining method. The branch-length estimation procedure is the same as above, except that the estimates among the six *E. caballus* sequences were taken from those of Fig. 2. This is because the branch-length estimation is generally more accurate when only closely related sequences are compared. A total of 97 nucleotide substitutions were estimated in the tree of Fig. 3. The root was located under the assumption of the molecular clock and the unrooted tree obtained by the neighbor-joining method. More specifically, branches of a pair of neighbors are successively averaged to find the root. For example, the height of the node connecting Thoroughbred 1 and 3 sequences is calcu-

Table 2. Estimated substitution events on the phylogenetic tree of Fig. 2

Nucleotide position[a]	Nucleotide change[b]
9	T(8) → C(2)
11	A(8) → G(2)
36	G(8) → A(7)
49	C(8) → T(2)
57	C(8) ↔ T(10)
95	T(7) → C(1)
102	G(8) → A(2), G(9) → A(5)
113	A(9) → G(4)
114	A(8) → G(7)
115	A(9) → G(5), A(10) → G(6)
121	T(8) → C(2)
122	G(8) → A(2)
153	A(9) → C(4), A(9) → T(5), C(8) ↔ A(10) [or C(10) → A(6), C(9) → T(5)]
167	A(8) → G(2)
168	A(10) → G(9)
184	G(10) → A(9)
221	T(9) → C(5)
255	T(8) ↔ C(10)
257	A(8) → T(2)
?	?(8) → ?(2)

[a] Nucleotide positions of Fig. 1. The last row with "?" indicates one unknown change inferred in the tree of Fig. 2

[b] Numbers in parentheses designate sequences and internal nodes as follows: 1 = T1, 2 = T2, 3 = T3, 4 = Pr, 5 = Mo, 6 = Jp, 7 = node connecting T1 and T3, 8 = node connecting T2 and node 7, 9 = node connecting Pr and Mo, and 10 = node connecting Jp and node 9. Species abbreviations follow those of Table 1. An arrow "→" indicates the direction of change, and a double arrow "↔" indicates ambiguity of direction. At position 153, two possibilities are presented. The last row indicates an extra substitution out of estimated nine at the branch going to the Thoroughbred 2 (T2) sequence

lated to be (1 + 0)/2 = 0.5. Then the height of the node connecting the former node and Thoroughbred 2 becomes [(2 + 0.5) + 9]/2 = 5.75. This averaging procedure is continued until the largest height (the root) is found. The root thus determined was located at 22.34 substitutions apart from *E. zebra*. Therefore, the point that divides the branch to *E. zebra* into 22 and 3 was determined as the root of the neighbor-joining tree (Fig. 3). This procedure is different from the usual midpoint rooting, in which only a pair of sequences that show the largest patristic distance is used. We also tried to use the bovine sequence (Anderson et al. 1982) as an outgroup to locate the root. However, the multiple alignment was not easy, and we did not incorporate the bovine sequence for comparison.

Numbers above the six internal branches are bootstrap probabilities (in %) based on 1,000 bootstrap trees. Six *Equus caballus* sequences are monophyletic with a high bootstrap probability (98%), and the clustering of *E. asinus* and *E. grevyi* is moderately supported (73% of bootstrap probability). Because of the inclusion of three other sequences, bootstrap probabilities within the six *Equus caballus* sequences were slightly lower than those of Fig. 2.

Table 3 presents the two kinds of estimated numbers of nucleotide substitution. Numbers below the diagonal are pairwise evolutionary distances computed by using Tamura and Nei's (1993) substitution model with a gamma distribution ($\alpha = 0.5$). These pairwise distances are larger than those when Kimura's (1980) method was used (data not shown). The increase was 5–14% within the six *E. caballus* sequences, while it was 18–48% among the *Equus* sequences. Those above the diagonal are patristic distances reconstructed from the tree of Fig. 3. Though some of those patristic distances are smaller than the corresponding original distances below the diagonal, both of them are larger than the nucleotide difference shown in Table 1. For example, the number of nucleotide differences between *E. zebra* and *E. grevyi* was 30 (21 transitional and 9 transversional differences; see Table 1), while the corresponding original and patristic distances are 41.5 and 43, respectively (Table 3). Because of these large differences, estimation of substitution events, involving different *Equus* species, was impossible.

Three equally parsimonious trees, requiring 81 nucleotide substitutions, were found by using the branch-and-bound option of PAUP. The topology of one of those, tree 1, was identical with that of the neighbor-joining tree (Fig. 3). Numbers in parentheses below the six internal branches of Fig. 3 are bootstrap probabilities (in %) based on 1,000 bootstrapped maximum parsimony trees. Those probabilities are slightly lower than those for the neighbor-joining trees except for one for the Thoroughbred 1 and 3 cluster (Fig. 3). The topological difference between the tree of Fig. 3 and the other two equally parsimonious trees (trees 2 and 3; not shown) resides in the position of the root for the six *Equus caballus* sequences; the root at the branch separating the cluster of Thoroughbred 1 and 3 and the remaining four sequences for tree 2, and that at the branch separating Thoroughbred 2 and the remaining five sequences for tree 3. Nucleotide positions 255, 36, and 9 supported trees 1, 2, and 3, respectively. Although the root (the position of the nearest common ancestor) for the six *Equus* sequences is not clear, the root is estimated to be always between some Thoroughbred horse(s) and the other horses.

Reanalysis of Restriction-Site Data

George and Ryder (1986) studied restriction-site variation of mtDNA for several *Equus* species. We reanalyzed their data in the following way. The maximum likelihood method of Nei and Tajima (1983) was used for estimating numbers of nucleotide substitution for all pairs of sequences, based on the restriction maps of Figs. 1 and 3 of George and Ryder (1986). Evolutionary distances thus obtained (Table 4) were not much different from those presented by George and Ryder (1986), who used Nei and Li's (1979) method. Figure 4 shows a neighbor-

Table 3. Estimated number of nucleotide substitutions among the nine equine mtDNA sequences[a]

Sequence name (abbr.)	T1	T2	T3	Pr	Mo	Jp	Zb	As	Gv
Thoroughbred 1 (T1)	—	13	1	10	12	7	43	39	36
Thoroughbred 2 (T2)	12.8	—	12	16	18	12	49	45	42
Thoroughbred 3 (T3)	1.0	11.7	—	9	11	6	42	38	35
Przewalskii wild (Pr)	9.1	15.6	8.0	—	6	5	45	41	38
Mongolian native (Mo)	12.9	16.7	11.7	5.4	—	7	47	43	40
Japanese native (Jp)	7.6	13.7	6.5	5.6	5.4	—	42	38	35
E. zebra (Zb)	43.8	45.4	41.8	49.0	46.5	43.7	—	46	43
E. asinus (As)	37.0	51.7	35.1	41.3	45.8	35.5	47.9	—	29
E. grevyi (Gv)	38.6	46.8	36.6	34.8	34.8	38.2	41.5	29.1	—

[a] Numbers below and above diagonal are original and patristic (under the tree of Fig. 3) number of nucleotide substitutions, respectively, within the 269 nucleotide positions compared

Table 4. Estimated numbers of nucleotide substitution per site among the nine mtDNA sequence: data of George and Ryder (1986) were used[a]

Eb	0.0725							
	(0.0135)							
Eg	0.0728	0.0341						
	(0.0132)	(0.0079)						
Ez	0.0743	0.0567	0.0500					
	(0.0138)	(0.0113)	(0.0101)					
Ea	0.0684	0.0640	0.0532	0.0615				
	(0.0126)	(0.0120)	(0.0104)	(0.0117)				
Eh	0.0818	0.0722	0.0683	0.0695	0.0639			
	(0.0147)	(0.0134)	(0.0126)	(0.0130)	(0.0120)			
Ec1	0.0045	0.0710	0.0713	0.0683	0.0669	0.0756		
	(0.0027)	(0.0133)	(0.0131)	(0.0129)	(0.0125)	(0.0140)		
Ec2	0.0029	0.0711	0.0714	0.0728	0.0671	0.0755	0.0045	
	(0.0021)	(0.0132)	(0.0129)	(0.0134)	(0.0124)	(0.0138)	(0.0026)	
Ec3	0.0044	0.0725	0.0728	0.0742	0.0685	0.0770	0.0059	0.0014
	(0.0026)	(0.0133)	(0.0131)	(0.0135)	(0.0125)	(0.0139)	(0.0030)	(0.0014)
	Ep	Eb	Eg	Ez	Ea	Eh	Ec1	Ec2

[a] Numbers in parentheses are standard errors. Ep = *Equus przewalskii* (=Prezewalskii's wild horse); Eb = *Equus burchelli antiquorum;* Eg = *Equus grevyi;* Ez = *Equus zebra hartmannae;* Ea = *Equus africanus somalicus* (=*Equus asinus*); Eh = *Equus hemionus,* Ec1 = Arabian horses; Ec2 = quarter and Peruvian paso horses; and Ec3 = Morgan horse

joining tree based on the distance matrix of Table 4. The rooting procedure is the same as that used in Fig. 3.

Sequence variation within the *E. caballus* (including the Przewalskii's wild horse) is much smaller than that between various *Equus* species. This confirms the result of George and Ryder (1986), and this situation is similar to the tree of Fig. 3. The topological relationship within the four horse sequences is identical with the result of George and Ryder (1986), while the branching pattern among the six equine species of Fig. 4 corresponds to tree c of Fig. 4 of George and Ryder (1986), which required one additional change compared to the two maximum parsimonious trees. Because the branch going to the cluster of *E. asinus* (=*E. africanus somalicus* in their nomenclature) and *E. hemionus* is quite short (see Fig. 4), tree c is not much different from trees a and b.

Four species (*E. caballus, E. asinus, E. zebra,* and *E. grevyi*) are shared between trees of Figs. 3 and 4, but the unrooted topological relationship is different from each other. *E. asinus* and *E. grevyi* are clustered in Fig. 3, while *E. zebra* and *E. grevyi* are clustered in Fig. 4. The bootstrap probability (73% and 59% for the neighbor-joining and the maximum parsimony trees, respectively) for the branch clustering *E. asinus* and *E. grevyi* is not high in Fig. 3, and the standard errors of evolutionary distances (Table 4) used to produce the tree of Fig. 4 are relatively large. Apparently, more sequence data are necessary to resolve the branching order among different *Equus* species.

Evolutionary Rates

The root of the phylogenetic tree shown in Fig. 3 was located on the assumption of the molecular clock, and a point within the branch going to *E. zebra* was chosen as the root. If we consider *E. zebra* as the outgroup, an approximate constancy seems to hold for the remaining part of the tree of Fig. 3. Therefore, we estimated the rate (λ) of nucleotide substitution using the equation $d = 2\lambda t$, where d is the number of nucleotide substitutions per site between a pair of sequences, and t is the divergence time. Higuchi et al. (1987) considered 2.0–3.3 million years

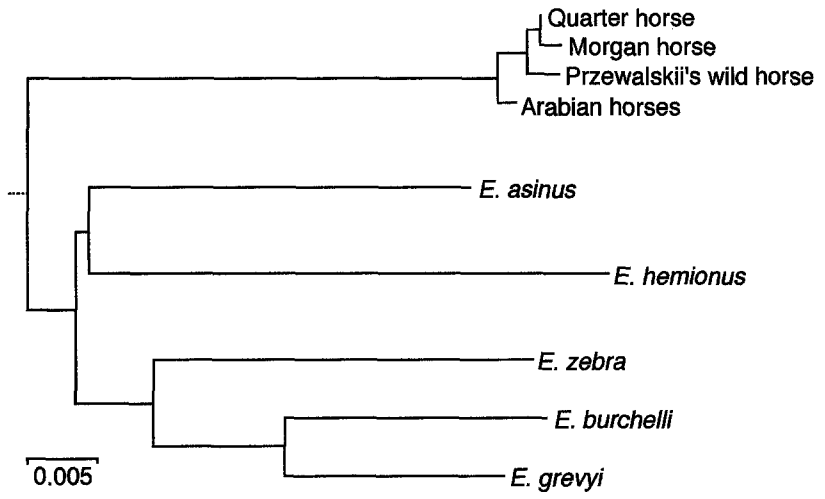

Fig. 4. A neighbor-joining tree of the nine *Equus* mtDNA sequences based on the evolutionary distance matrix of Table 4. Original data were from George and Ryder (1986). *E. asinus* and Przewalskii's wild horse are *E. africanus somalicus* and *E. przewalskii* in their nomenclature.

for the divergence between horse (*E. caballus*) and zebras (*E. zebra* and *E. burchelli*). In our tree of Fig. 3, however, the lineage of *E. zebra* was considered to diverge first. Lindsay et al. (1980) suggested that the ancestral *Equus* species migrated from North America to Eurasia at around 2.5–3.0 million years ago, and MacFadden (1992) wrote that "*Equus* was essentially monospecific in both North American and Eurasia during the Pliocene." In any case, there is always uncertainty on the correspondence between the fossil record and the molecular data. Therefore, we used a wider, more conservative range of divergence time. The time (t) of the root of the tree of Fig. 3 was assumed to correspond to 2–4 million years ago. Because the largest divergence had a mean evolutionary distance of 44 nucleotide substitutions among the 269 sites (see Fig. 3), $d = 0.164$ (=44/269). Thus the rate ($\lambda = d/2t$) becomes approximately $2–4 \times 10^{-8}$ per nucleotide site per year.

Higuchi et al. (1987) found about 5–7% difference between horse and zebras for the coding regions of cytochrome oxidase I and NADH dehydrogenase I genes. If we assume the same divergence time estimate (2–4 million years ago) used in the present study, the evolutionary rate for that region becomes approximately $0.6–1.8 \times 10^{-8}$ per nucleotide site per year, lower than our estimate for the D-loop region. (Note that the definition of this evolutionary rate is different from Higuchi et al.'s (1987) "the mean rate of divergence,"—that is, twice this rate.) This difference occurred apparently because of the functional differences of the coding and noncoding sequences of mtDNA, and this supports the notion that the noncoding region mtDNA is more variable than its coding sequences (Aquadro and Greenberg 1982; Cann et al. 1984).

Reanalysis of the restriction-site data of George and Ryder (1986) gives the estimate for the largest divergence to be 0.072 (Fig. 4). Therefore, the evolutionary rate becomes $0.9–1.8 \times 10^{-8}$ per nucleotide site per year, assuming the divergence time range adopted in the present study. This rate is again lower than that for the D-loop region.

Tamura and Nei (1993) estimated the rate of nucleotide substitution for the entire control (D-loop) region of human mtDNA to be $2.5–15 \times 10^{-8}$ per site per year (95% lower and upper bounds). Our estimate for equine species ($2–4 \times 10^{-8}$) overlaps with that for human, but it is smaller than the modal rate (7.5×10^{-8}) for human (Tamura and Nei 1993). Because we used only a ca. 270-bp variable region out of the 1,100-bp control region found by Ishida et al. (1994), it is possible that the substitution rate for the entire D-loop region of the equine mtDNA may be much smaller than that for the human mtDNA.

Under the tree of Fig. 3, we can map six insertions or deletions on the 0.27-kb sequence data. (Gaps of *E. zebra* and *E. grevyi* at position 47 [see Fig. 1] are considered to occur independently according to Fig. 3.) Because the total number of nucleotide substitutions was 97 for that tree, it corresponds to 8.8–17.6 million years (=[97/22]/[2–4 million years]). Therefore, if we use method 1 of Saitou and Ueda (1994), the rate of insertion and deletion becomes 1.3–2.5 per kilobase per million year (=[6/0.27 kb]/[8.8–17.6 million years]). Interestingly, this is in good agreement with the corresponding rate (ca. 2.0 per kilobase per million year) for the primate noncoding mtDNA (Saitou and Ueda 1994).

Emergence of the Przewalskii's Wild Horse Population

Finally, we would like to discuss the evolutionary history of the Przewalskii's wild horse population. It is noteworthy that the lineage of the Przewalskii's wild horse is not located at the deepest branching among the *E. caballus* sequences in either Fig. 3 or 4. This pattern contradicts with the hypothesis that the domestic horse was derived from the Przewalskii's wild horse lineage. In this context, Benirschke et al.'s (1965) hypothesis on the evolution of horse karyotype (see Introduction) should be re-

considered. If we accept the branching pattern of Figs. 3 and 4 regarding the position of the Przewalskii's wild horse, its chromosome type (2n = 66) seems to be derived from the chromosome type (2n = 64) of the domestic horse.

Of course, a final clarification of this point waits for more accumulation of genetic polymorphism data of *Equus caballus*. If this is the case, however, when did this chromosomal change occur? The largest divergence within the six *E. caballus* sequences had a mean evolutionary distance of 11.5 nucleotide substitutions among the 271 sites (Fig. 2), and the $d = 0.042$ (=11.5/271). Under the assumption of the molecular clock, the corresponding divergence time ($t = d/2\lambda$) becomes approximately 0.5–1 million years (=0.042/ [4–8 × 10^{-8}]). It is interesting to note that this "coalescence time" for *E. caballus* seems to be somewhat larger than that for human but comparable to that for common chimpanzee (Tamura and Nei 1993). In any case, emergence of the Przewalskii's wild horse population should be much later than this coalescence, according to the topology of the tree of Fig. 2. Therefore, the change of chromosome type on the lineage of the Przewalskii's wild horse, probably the increase of the chromosome number from 64 to 66, seemed to happen within a half-million years. Such a swift change of chromosome number is in fact a common feature for the genus *Equus* (Bush et al. 1977).

The above argument solely relies on the genealogy of mtDNA. However, that may be different from the genealogy of nuclear DNA, as observed in Japanese wild mice (Yonekawa et al. 1988). Przewalskii's wild horse was clustered with the Mongolian native horse in the tree of Fig. 2. Thus it is not excluded that a certain degree of gene flow occurred between the ancestral populations of the Przewalskii's wild horse and the Mongolian native horse. In fact, a past distribution of the Przewalskii's wild horse overlapped with the area where the Mongolian native horses are distributed now (Mohr 1971; Bökönyiu 1974). It also should be mentioned that the Japanese native horses are considered to have derived from the Mongolian native horses and have been kept as breeds (Nozawa 1992). Therefore, detailed examination of nuclear DNA at the molecular level is necessary to access our hypothesis on the emergence of the Przewalskii's wild horse population.

Acknowledgments. In collecting blood samples, we thank Masao Saito of Tama Zoological Park in Tokyo, Yasunori Iwamoto of Yumemigasaki Zoological Park in Kawasaki, and Eigi Taneda of The Equestrian Park, JRA, in Tokyo. This study was partially supported by a Grant-in-Aid for Scientific Research on Priority Areas (Molecular Evolution) of Ministry of Education, Science and Culture (Japan) to N.S.

References

Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717

Aquadro CF, Greenberg BD (1982) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103:287–312

Benirschke K, Malouf N, Low RJ, Heck H (1965) Chromosome complement differences between *Equus* caballus and *Equus* przewalskii, Poliakoff. Science 148:382–383

Bökönyi S (1974) History of domestic mammals in Central and Eastern Europe. Akadémiai Kiadó, Budapest, 230–296

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76:1967–1971

Bush GL, Case SM, Wilson AC, Patton JL (1977) Rapid speciation and chromosomal evolution in mammals. Proc Natl Acad Sci USA 74:3942–3946

Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 325:31–36

Chang D, Clayton DA (1985) Priming of human mitochondrial DNA replication occurs at the light strand promoter. Proc Natl Acad Sci USA 82:351–355

Clegg JB (1974) Horse hemoglobin polymorphism. Ann NY Acad Sci 241:61–69

Fitch WM (1977) On the problem of discovering the most parsimonious tree. Am Naturalist 111:223–257

George M Jr, Ryder OA (1986) Mitochondrial DNA evolution in the genus *Equus*. Mol Biol Evol 3:535–546

Groves CP, Willoughby DP (1981) Studies on the taxonomy and phylogeny of the genus *Equus*. 1. Subgeneric classification of the recent species. Mammalia 45:321–354

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Higuchi RG Wrischnik LA, Oakes E, George M, Tong B, Wilson AC (1987) Mitochondrial DNA of the extinct quagga: relatedness and extent of postmortem change. J Mol Evol 25:283–287

Ishida N, Hasegawa T, Takeda K, Sakagami M, Onishi A, Inumaru S, Komatsu M, Mukoyama H (1994) Polymorphic sequence in the D-loop region of equine mitochondrial DNA. Anim Genet 25:215–221

Kaminski M (1979) The biochemical evolution of the horse. Comput Biochem Physiol 63B:175–178

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park, PA

Lindsay EH, Opdyke ND, Johnson ND (1980) Pliocene dispersal of the horse *Equus* and late Cenozoic mammalian dispersal events. Nature 287:135–138

MacFadden BJ (1992) Fossil horses: systematics, paleobiology, and evolution of the family *Equidae*. Cambridge University Press, Cambridge

Mohr E (1971) The Asiatic wild horse. JA Allen, London, pp 26–33. Translated by Goodall DM from "Das Urwildpferd," A Zaisen Verlag, Wittenberg Lutherstadt, 1970

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. Genetics 105: 207–217

Nerurkar VR, Song KJ, Saitou N, Mallan RR, Yanagihara R (1993) Interfamilial and intrafamilial genomic diversity of human T lymphotropic virus type I strains from Papua New Guinea and the Solomon Islands. Virology 196:506–513

Nowak RM, Paradiso JL (1983) Walker's mammals of the world, 4th edition. The Johns Hopkins University Press, Baltimore

Nozawa K (1992) Origin and ancestry of native horses in Eastern Asia and Japan. Jpn J Equine Sci 3:1–18

Ryder OA, Epel NC, Benirschke K (1978) Chromosome banding studies of the *Equidae*. Cytogenet Cell Genet 20:323–350

Ryder OA, Starkes RS, Starkes MC, Clegg JB (1979) Hemoglobin polymorphism in *Equus przewalskii* and *E. caballus* analyzed by isoelectric focusing. Comput Biochem Physiol 62B:305–308

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Saitou N, Ueda S (1994) Evolutionary rate of insertions and deletions in non-coding nucleotide sequences of primates. Mol Biol Evol 11:504–522

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 9.14–9.23

Sanger F, Niklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Simpson GG (1951) Horses: the story of the horse family in the modern world and through sixty million years of history. Oxford University Press, New York

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1.1. Computer program distributed by the Illinois Natural History Survey, Champaign, IL

Tamura K and Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Trommerhausen-Smith A, Ryder OA, Suzuki Y (1979) Bloodtyping studies of twelve przewalskii's horses. Intl Zoo Yrbk 9:224–227

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421

Yonekawa H, Moriwaki K, Gotoh O, Miyashita N, Matsushima Y, Shi L, Cho WS, Zehn XL, Tagashira Y (1988) Hybrid origin of Japanese mice "*Mus musculus* molossinus": evidence from restriction analysis of mitochondrial DNA. Mol Biol Evol 5:63–78