J Mol Evol (1996) 43:438–446

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1996

The Complete Mitochondrial DNA (mtDNA) of the Donkey and mtDNA Comparisons Among Four Closely Related Mammalian Species-Pairs

Xiufeng Xu, Anette Gullberg, Ulfur Arnason

Division of Evolutionary Molecular Systematics, University of Lund, Sölvegatan 29, S-223 62 Lund, Sweden

Received: 15 October 1995 / Accepted: 15 April 1996

Abstract. The nucleotide sequence of the complete mitochondrial genome of the donkey, *Equus asinus*, was determined. The length of the molecule is 16,670 bp. The length, however, is not absolute due to pronounced heteroplasmy caused by variable numbers of two types of repetitive motifs in the control region. The sequence of the repeats is (a) 5′-CACACCCA and (b) 5′-TGCGCGCA, respectively. The order of (a) and (b) can be expressed as {*n*[2(a)+(b)]+*m*(a)}. In 32 different clones analyzed the number of *n* and *m* ranged from 0 to 9 and 1 to 7. The two rRNA genes, the 13 peptide-coding genes, and the 22 tRNA genes of the donkey and the horse, *Equus caballus*, were compared in detail. Total nucleotide difference outside the control region was 6.9%. Nucleotide difference between peptide-coding genes ranged from 6.4% to 9.4% with a mean of 8.0%. In the inferred protein sequences of the 13 peptide-coding genes the amino acid difference was 0.2–8.8%, and the mean for the 13 concatenated amino acid sequences was 1.9%. In the 22 tRNA genes, the mean difference was 3.5%, and that in the two rRNA genes was 4.1%. The mtDNA differences between the donkey and the horse suggest that the evolutionary separation of the two species occurred ≈9 million years ago. Analyses of differences among the mtDNAs of three other species-pairs, harbor seal/grey seal, fin whale/blue whale, and *Homo*/common chimpanzee, showed that the relative evolutionary rate of individual peptide-coding genes varies among different species-pairs and modes of comparison. The findings show that the superimposition of sequence data of one lineage for resolving and dating evolutionary divergences of other lineages should be performed with caution unless based on comprehensive data.

Key words: Mitochondrial DNA — Molecular comparisons — Molecular phylogeny — Mammalian species-pairs — Equidae — Donkey — Horse

Introduction

During the last few years there has been a considerable progress in the sequencing of complete mammalian mtDNA molecules. There are, however, only four orders represented by two or more complete mtDNAs. These orders are Primates, represented by *Homo* (Anderson et al. 1981; Ozawa et al. 1991; Horai et al. 1995; Arnason et al. 1996a), common chimpanzee (Horai et al. 1995; Arnason et al. 1996a), pygmy chimpanzee (Horai et al. 1995), gorilla (Horai et al. 1995; Xu and Arnason 1996), and Bornean orangutan (Horai et al. 1995); Rodentia, represented by the mouse (Bibb et al. 1981) and the brown rat (Gadaleta et al. 1989); Cetacea, represented by the fin (Arnason et al. 1991a) and blue (Arnason and Gullberg 1993) whales; and Carnivora, represented by the harbor (Arnason and Johnsson 1992) and grey seals (Arnason et al. 1993). Mammalian orders represented by a single complete mtDNA molecule are Artiodactyla, represented by the cow (Anderson et al. 1982); Perissodactyla, represented by the horse (Xu and Arnason 1994); Marsupialia, represented by the opossum (Janke et al. 1994); Lipotyphla, represented by the hedgehog

Correspondence to: U. Arnason

(Krettek et al. 1995); and Monotremata, represented by the platypus (Janke et al. 1996). While comparisons of distantly related mtDNAs have particular interest for establishing long-range evolutionary relationships, comparisons among mtDNAs of closely related species are valuable for studies of molecular evolution among sequences that are unaffected by molecular saturation and for establishing the degree of molecular difference between species that may still produce offspring. The latter issue has been dealt with in some detail in analyses of the fin and blue whales, which, despite their molecular distinction, have been shown to produce offspring where the females are not obligatorily sterile (Arnason et al. 1991b; Spilliaert et al. 1991).

In the present study we compare the complete mtDNA molecules of two mammalian species, the donkey and the horse, which also can produce offspring. The offspring is sterile, however, and is produced only under domestication. The family Equidae, horses, comprises a single genus, *Equus,* with nine extant species. The family is well defined and is distinct from the other two families of the order Perissodactyla—namely, the Rhinocerotidae and the Tapiridae. Based on classical approaches it has been postulated that members of the genus *Equus* diverged during the last 3–5 Myr (million years) (Simpson 1951; Lindsay et al. 1980). The differences among various equiid mtDNAs have been estimated on the basis of restriction-endonculease analyses (George and Ryder 1986). Using a divergence rate of 2% per million years, the authors suggested a primary evolutionary divergence of equiid lineages about 3.9 MYA (million years ago).

In the present study we report the complete nucleotide sequence of the mitochondrial genome of the donkey and propose a dating of the divergence between the donkey and the horse. In addition to the comparison between these two species we also compare each of the 13 peptide-coding genes and the 12S and 16S rRNA genes among three additional closely related species-pairs, the harbor and the grey seals, the fin and the blue whales, and *Homo* and the common chimpanzee. This comparison of the peptide-coding genes was undertaken because it has been shown previously (Arnason and Johnsson 1992; Cao et al. 1994) that individual mtDNA genes may provide different topologies for ordinal relationships among mammals. In the present paper we address this issue from another angle by comparing the differences between individual peptide-coding genes among four closely related mammalian species-pairs and by studying to what extent differences in one gene of one species-pair are paralleled by similar differences in the same gene of other species-pairs.

Materials and Methods

Mitochondrial DNA was isolated from the frozen kidney of a healthy adult donkey following the procedure described by Arnason et al. (1991a). The tissue was provided by Dr. Stuart W. J. Reid, University of Glasgow Veterinary School, Glasgow, UK. The purified mtDNA was digested separately with *Hin*dIII, *Bln*I, *Spe*I, *Xba*I, and *Bam*HI and the fragments were separated on a preparative agarose gel. The fragments were excised from the gel and extracted by electroelution. Fragments shorter than 4,000 nt (nucleotides) were ligated into M13 vectors, whereas larger fragments of *Xba*I, *Spe*I, and *Bam*HI were redigested with enzymes producing compatible ends and thereafter ligated into M13. Positive clones were identified by hybridization using mtDNA of the horse as labeled probes.

Sequencing was according to the dideoxy chain termination technique with [^{35}S]dATP. The work was performed manually using both universal and a number of specific oligodeoxynucleotide primers. In addition to natural (not PCR) clones the repetitive portion of the control region was also sequenced after M13 cloning of PCR-amplified fragments. The sequence of each clone was determined on the basis of three independent sequencing reactions.

The mtDNA sequence of the donkey has been deposited at EMBL with accession number X97337. Users of the sequence are kindly requested to refer to the present paper and not only to the accession number of the sequence.

Results

General Features of the Mitochondrial Genome of the Donkey

The length of the mtDNA molecule of the donkey, *Equus asinus,* presented here is 16,670 nt. This length is not absolute, however, because of variation in the number of two kinds of repetitive motifs in the control region of the molecule. The 16,670-nt variety of the molecule includes 22 copies of one motif (a) and nine copies of the other (b) (see below for characterization, organization, and number of motifs). The base composition of the L-strand is A, 32.4%; C, 28.9%; G, 13.2%; T, 25.5%. The organization of the molecule is the same as that of other eutherian mtDNAs described.

With the exception of the NADH5 (nicotinamide adenine dinucleotide dehydrogenase subunit 5) gene all peptide-coding genes of the donkey have a methionine start codon. The start codon of NADH5 is ATT (isoleucine). ATT is not an uncommon start codon, however, in mammalian mtDNAs. In the horse it is also isoleucine (ATC) and in the harbor and the grey seals it is ATT. Three of the 13 peptide-coding genes, COIII (cytochrome c oxidase subunit III), NADH3, and NADH4, are not terminated by a complete stop codon. A stop codon is absent in the COIII gene of all mammalian mtDNAs described so far, except in the fin and the blue whales (e.g., Table II in Xu and Arnason 1994). The absence of a complete stop codon in the NADH4 gene is a common characteristic of mammalian mtDNAs and the NADH3 gene is terminated by a stop codon only in the mouse and the rat. Like the horse, these genes of the donkey are terminated with an incomplete stop codon, T or TA. Termination of this kind is not an uncommon feature of mammalian mtDNA peptide-coding genes (Wolstenholme 1992; Janke et al. 1994; Xu and Arnason 1994),

Table 1. Number of repetitive motifs in 32 clones of the mitochondrial control region of the donkey[a]

Clone	*n*	*m*	Clone	*n*	*m*	Clone	*n*	*m*	Clone	*n*	*m*
1	0	1	9	2	2	17	4	4	25	7	5
2	0	2	10	2	3	18	4	5	26	7	7
3	0	6	11	2	3	19	6	4	27	8	3
4	1	1	12	3	2	20	6	4	28	8	4
5	1	3	13	3	3	21	7	3	29	8	4
6	1	3	14	3	4	22	7	4	30	8	5
7	2	1	15	3	4	23	7	4	31	8	6
8	2	1	16	3	5	24	7	4	32	9	4

[a] The repetitive sequence contains two different motifs, (a) 5′-CACACCCA and (b) 5′-TGCGCGCA. The motifs are arranged as {*n*[2(a)+(b)]+*m*(a)}

consistent with the notion (Ojala et al. 1981) that the transcripts of such genes contain a stop codon created by post-transcriptional polyadenylation.

The control region of the mtDNA presently described is 1,207 nt long. The region is characterized by two different repeat motifs: (a) 5′-CACACCCA and (b) 5′-TGCGCGCA. No variation was observed within the sequence of each motif. The latter motif has the pyrimidine/purine alternation found in most tandemly organized repeat motifs occurring in mitochondrial control regions. Examination of 12 natural and 20 PCR clones revealed pronounced heteroplasmy among the different mtDNA molecules. Among the 32 clones 24 different types of motif arrangements were identified. The arrangement of the repeats can be expressed: {*n*[2(a)+(b)]+*m*(a)}, where *n* can run from 0 to 9 and *m* from 1 to 7. Table 1 gives the numbers for *n* and *m* among the 32 examined clones. The clones are numbered according to increasing values of *n* and *m*. Motif (a) was absent in three clones. In the presently described complete sequence $n = 9$ and $m = 4$.

Excluding tRNA-Ser(UCN) and tRNA-Ser(AGY), the inferred secondary structures of the remaining 20 tRNA genes of the donkey conform with the common structures of mitochondrial tRNA genes described by Kumazawa and Nishida (1993). The tRNA stems AA (amino acid-acceptor), D (dihydrouridine), AC (anticodon), and T (TΨC) are 7, 4, 5, and 5 base pairs long, respectively. The regions separating stems AA and D, D and AC, AC and T, and T and AA are 2, 1, 3–5, and 0 nt long, respectively. In the two deviating tRNA-Ser genes, tRNA-Ser(AGY) has no D stem and the AC stem in tRNA-Ser(UCN) is 6 base pairs long. Furthermore, the junction of the AA-D stem in the tRNA-Ser(UCN) has only 1 nt.

Comparison Between the mtDNAs of the Donkey and the Horse

The percentage base compositions of the mtDNAs of donkey/horse are very similar: A, 34.2/32.2; C, 28.9/28.5; G, 13.2/13.4; T, 25.5/25.9. Outside the control regions the mtDNAs of the donkey and the horse differ by 6.9%. An alignment of the two sequences, excluding control regions, shows 14 indels (insertions or deletions). Six of these occur in the rRNA genes and eight in the tRNA genes, respectively. The stem portions of the origin of L-strand replication, which are highly conserved among mammals, are identical in the donkey and the horse. The loops differ by a single transition.

We will provide separate accounts of the comparisons between the control region, the rRNA genes, the tRNA genes, and the peptide-coding genes of the mtDNAs of the donkey and the horse.

The Control Region

The length of the control region of the mtDNA of the donkey, excluding the repetitive portion, is 959 nt, one less than that of the horse. There are 108 differences (11.2%) in an alignment of the two regions, 88 transitions, 16 transversions, and four indels. The repetitive motifs occur in the same region in both species. The length of the repeated motif of the horse, 5′-GTGCACCT, is the same as that of the two different donkey motifs, but the sequence of each of the three motifs is highly distinct.

The Peptide-Coding Genes

The lengths of each of the 13 mitochondrial peptide-coding genes are the same in the donkey and the horse. The results of a comparison between individual genes of the two species are shown in Table 2. Besides giving the number of nt differences, the number of amino acid (aa) differences, and the number of conservative nt differences, the table also details differences with respect to codon position and type (transition or transversion). Conservative nt differences (Irwin et al. 1991) constitute (a) all substitutions in codon position 1, except synonymous leucine transitions, (b) all substitutions in codon position 2, and (c) all transversions in codon position 3.

Total nucleotide difference between peptide-coding genes of the donkey and the horse ranges from 6.4% in NADH3 to 9.4% in NADH5. The mean nt difference is 8.0%. It should be observed that this figure is the mean for the combined length of all peptide-coding genes (11,403 nt), and not the mean of the values for individual genes. Conservative differences range from 0.4% in COI and COIII to 4.4% in ATPase8 with a mean of 1.2%. The aa differences range from 0.2% in COI to 8.8% in ATPase8 among H-strand-encoded genes. The aa difference for the L-strand-encoded NADH6 gene is 5.7%; the conservative nt difference for this gene is also high, 2.7%. The mean aa differences for the combined length of the peptide-coding genes is 1.9%.

The total number of nt substitutions in the 13 peptide-

Table 2. Nucleotide difference with respect to codon position (1,2,3) between each of the 13 mitochondrial peptide-coding genes of the donkey and horse, and aa differences in the same genes[a]

		1					3				
		Ti			2		Ti		Tv		
Gene	Length	a	b	Tv	Ti	Tv	a	b	a	b	aa diff.
NADH1	957	7	3	–	1	–	9	42	5	1	5
NADH2	1,041	10	9	–	7	–	8	55	1	6	13
COI	1,545	3	1	–	–	–	16	84	–	5	1
COII	684	9	1	–	–	–	2	34	–	3	1
ATPase8	204	–	3	–	4	–	–	7	–	2	6
ATPase6	681	1	5	–	2	–	10	41	–	2	7
COIII	783	4	–	1	–	–	6	42	–	2	1
NADH3	345	2	1	–	–	–	6	11	–	2	1
NADH4L	297	3	1	–	1	–	2	13	–	2	2
NADH4	1,377	6	2	1	4	–	13	81	4	3	7
NADH5	1,821	10	7	1	6	1	22	111	4	9	17
NADH6	528	1	6	1	3	1	4	16	–	3	10
Cytb	1,140	4	3	–	–	–	10	62	2	7	3
Total	11,403	60	42	4	28	2	108	599	16	47	74
Conservative diff.			46		30				63		
Total Diff.			106		30			770			
Ratio total diff.			3.5		1.0			25.7			
Ratio conservative diff.			1.5		1.0			2.1			

[a] Ti: transitions; Tv: transversions; a: differences involving leucine in both species, b: differences other than those involving leucine

coding genes is 906, including 139 conservative nt substitutions. The ratio for all nt substitutions according to codon position is 3.5 : 1.0 : 25.7. The corresponding ratio for conservative nt substitutions is 1.5 : 1.0 : 2.1.

The rRNA Genes

Between the 12S rRNA genes of the donkey and the horse there are 48 differences (4.9%), 38 transitions (3.9%), seven transversions (0.7%), and three indels (0.3%). The corresponding values for the 16S rRNA genes are 58 (3.7%), 46 (2.9%), nine (0.6%), and three (0.2%). The differences for the combined lengths of the two genes are: total difference 4.1%, transition difference 3.3%, transversion difference 0.6%, and indels 0.2%.

The tRNA Genes

The combined length of the tRNA genes of the donkey is 1,517 nt as compared with 1,520 nt in the horse. The results of a comparison between the individual tRNA genes are shown in Table 3. Five genes are identical in the two species. The tRNA-Ser(AGY) and tRNA-Lys genes differ markedly between the two species with seven and five differences, respectively. The total number of differences in the tRNA genes is 54 (41 transitions, five transversions, and eight indels). It is noteworthy that all nucleotide differences (13) of the inferred stem regions were transitions and that all indel differences were limited to the loop regions (D loop and/or T loop).

The total sequence difference between the 22 tRNA genes of the donkey and the horse is 3.5%, as compared with 6.9% in the complete alignment outside the control region. The total number of differences in the stem regions of the 22 tRNA genes is only 13 (0.9%).

Comparison Among Closely Related Pairs of Mammalian Species

In order to study the evolution of individual peptide-coding genes in close evolutionary relationships the comparison between the donkey and the horse (order Perissodactyla) was extended by including closely related species-pairs of three additional mammalian orders, Carnivora (harbor and grey seals), Cetacea (fin and blue whales), and Primates (*Homo* and common chimpanzee). The results of the comparisons among the four species-pairs are summarized in Tables 4–6.

The order of the four species-pairs, arranged with respect to molecular difference, is: harbor/grey seals < horse/donkey < fin/blue whales < *Homo*/common chimpanzee. Table 4 shows the differences and the rank (in parenthesis) of the 13 peptide-coding genes when they are ranked according to increasing differences in each

Table 3. Lengths and number of nucleotide differences between the mitochondrial tRNAs of the donkey and horse[a]

	Length (nt) donkey/horse	Nucleotide difference			Nucleotide differences in stems	
		T	Tv	Gaps	Number	Region
tRNA-Phe	71/70	2		1		
tRNA-Val	67/67	2			1 (Ti)	T
tRNA-Leu (UUR)	74/75	2		1		
tRNA-Ile	69/69	2				
tRNA-Gln	73/73	2			1 (Ti)	AA
tRNA-Met	69/69					
tRNA-Trp	69/70	2	1	1	1 (Ti)	AC
tRNA-Ala	69/69	2			1 (Ti)	T
tRNA-Asn	73/73	2				
tRNA-Cys	66/68	3		2	1 (Ti)	AA
tRNA-Tyr	67/67					
tRNA-Ser (UCN)	69/69					
tRNA-Asp	67/67	2			1 (Ti)	AC
tRNA-Lys	69/68	5	2	1	1 (Ti)	D
tRNA-Gly	70/69	1		1		
tRNA-Arg	69/69					
tRNA-His	69/69	2				
tRNA-Ser (AGY)	60/60	7			3 (Ti)	T
tRNA-Leu (CUN)	70/70	2			2 (Ti)	AC; T
tRNA-Glu	69/69	1				
tRNA-Thr	72/74	2	2	1	1 (Ti)	T
tRNA-Pro	66/66					
Total	1,517/1,520	41	5	8	13 (Ti)	

[a] Ti: transition; Tv: transversion; AA: amino acid-acceptor stem; AC: anticodon stem; D: dihydrouridine stem; T: TΨC stem

pairwise comparison. As is apparent in Table 4, the same gene may exhibit quite different relative rates of evolution in different species-pairs. Thus, in the donkey/horse comparison the NADH3 gene shows the lowest total nt difference, 6.4%. The same gene has the second-lowest difference in the comparison between the seals, 2.3%. Between the two whales the difference is 8.7% (eighth ranking position), whereas between *Homo* and the chimpanzee the difference is 11.9% (13th position). Also among different modes of comparison pronounced differences may occur. Thus the total nt difference in ATPase8 is low, whereas it is very high in other modes of comparison.

The fact that the same peptide-coding gene shows different evolutionary rates in different pairwise comparisons makes it difficult to establish a general evolutionary rate for each gene. The conformity of the ranking of the 13 peptide-coding genes among the four pairwise comparisons was examined by first calculating the Kendall rank correlation coefficient τ (tau) for each of the six possible combinations among the four species-pairs and then forming the average $\bar{\tau}$ for the six τ coefficients. In addition to computing τ, we also calculated $\pi = (1 + \tau)/2$, a number between 0 and 1 which can be interpreted as the probability that two genes, chosen at random, will be ranked in the same order in the two comparisons. In the same way $\bar{\pi} = (1 + \bar{\tau})/2$ can be thought of as the probability that two genes, chosen at random, will be ranked in the same order in two comparisons, chosen at random from the four comparisons under study. For total nt substitution the mean rank correlation coefficient is 0.20 ($\bar{\pi} = 0.60$); for conservative nt substitution 0.39 ($\bar{\pi} = 0.70$); for total aa difference 0.38 ($\bar{\pi} = 0.69$); and for difference according to chemical properties of aa, 0.41 ($\bar{\pi} = 0.71$).

In order to assess the consistency among the four different modes of comparison and to establish the relative evolutionary rate of each peptide-coding gene, the ranking values of each gene (Table 4) were combined (Table 5). The order of the genes in this table, as in Table 4, is according to increasing difference based on conservative nt substitution, total aa difference, and difference according to chemical properties of amino acids. Each value in Table 5 constitutes the mean of the four pairwise comparisons in Table 4. As shown in Table 4 the total nt difference between the COI gene of the donkey and the horse is 7.1%. This gives the gene ranking position 4 (fourth lowest difference) in the horse/donkey comparison. For the same gene the difference between the two seals is 3.6% (ranking position 6); that between the two whales is 7.4% (ranking position 3); and that between *Homo* and the common chimpanzee is 8.8% (ranking position 4). Thus the mean ranking position for the COI with respect to total nt difference is 4.3. This is the second lowest ranking position for this type of comparison; only the NADH1 gene has a lower mean ranking position, 3.8. Table 5 shows the mean ranking values for each of the four modes of comparison—total nt substi-

Table 4. Percent difference between the mitochondrial peptide-coding genes of donkey (*E.as*)/horse (*E.ca*), harbor (*P.vi)*/grey (*H.gr*) seals, fin (*B.ph*)/blue (*B.mu*) whales, and *Homo*/chimpanzee (*P.tr*)[a]

	Nucleotide difference—total/conservative			
Gene	*E.as/E.ca*	*P.vi/H.gr*	*B.ph/B.mu*	*Homo/P.tr*
COI	(4) 7.1/0.4 (1)	(6) 3.6/0.5 (6)	(3) 7.4/0.6 (1)	(4) 8.8/0.8 (1)
COIII	(3) 7.0/0.4 (2)	(8) 4.0/0.4 (5)	(4) 7.5/0.9 (5)	(7) 9.6/1.4 (4)
COII	(6) 7.2/0.6 (3)	(7) 3.9/0.1 (2)	(6) 8.3/0.7 (4)	(6) 9.5/1.0 (3)
NADH3	(1) 6.4/0.9 (4)	(2) 2.3/0.3 (3)	(8) 8.7/0.6 (2)	(13) 11.9/2.9 (11)
NADH1	(5) 7.1/1.0 (6)	(4) 3.0/0.3 (4)	(1) 7.1/0.7 (3)	(5) 9.2/2.0 (7)
NADH4L	(7) 7.4/1.4 (9)	(1) 2.0/0.0 (1)	(10) 9.1/2.7 (12)	(1) 7.4/1.0 (2)
NADH4	(10) 8.3/1.0 (5)	(12) 4.6/0.9 (9)	(13) 10.5/1.7 (7)	(9) 9.7/1.8 (6)
Cytb	(8) 7.7/1.1 (7)	(9) 4.3/1.0 (10)	(5) 7.6/1.5 (6)	(12) 11.7/3.4 (13)
NADH2	(12) 9.2/2.2 (11)	(10) 4.4/0.8 (7)	(12) 9.6/2.6 (10)	(8) 9.6/1.8 (5)
NADH5	(13) 9.4/1.5 (10)	(11) 4.4/0.9 (8)	(7) 8.4/2.0 (8)	(11) 10.6/3.4 (12)
ATPase6	(11) 9.0/1.3 (8)	(5) 3.4/1.2 (12)	(11) 9.4/2.3 (9)	(2) 8.5/2.1 (8)
NADH6	(2) 6.6/2.7 (12)	(13) 6.1/1.1 (11)	(9) 8.9/2.7 (11)	(10) 9.7/2.1 (9)
ATPase8	(9) 7.8/4.4 (13)	(3) 2.9/1.5 (13)	(2) 7.3/3.1 (13)	(3) 8.7/2.4 (10)
Mean	8.0/1.2	4.0/0.7	8.5/1.5	9.8/2.1

[a] Difference with respect to chemical properties of amino acids is based on Gribskov and Burgess (1986). Figures in parentheses show the position with respect to increasing difference in each mode of comparison. The order of genes is according to increasing difference based on the conservative nucleotide substitution, total amino acid difference, and difference according to chemical properties of amino acids

tution, conservative nt substitution, total aa difference, and aa difference with respect to chemical properties. The Kendall rank correlation coefficient among the ranking of conservative nt difference, total aa difference, and aa difference with respect to chemical properties was 0.88, 0.81, and 0.86, respectively, with an average of 0.85 ($\overline{\pi} = 0.93$). Kendall's correlation coefficient for the ranking based on total nt difference compared with the ranking of the three other modes of comparison was 0.24 ($\overline{\pi} = 0.62$).

As is evident in Table 5, the ranking based on total nt difference differs considerably from that of the three other modes of comparison. This is also reflected in the rather low Kendall correlation coefficient (0.24) between the ranking of total nt difference and other differences. The most conspicuous deviation is the low ranking position (2/3) of ATPase 8 with respect to total nt difference and the high ranking position (13) of this gene in the other three comparisons. The situation is the reverse in NADH4, which has ranking position 13 with respect to total nucleotide difference as compared with 7 in the three other comparisons. The ranking position of the cytochrome *b* gene, frequently used in phylogenetic comparisons among mammals (Irwin et al. 1991; Irwin and Arnason 1994; Ma et al. 1993; Arnason et al. 1995; Arnason and Gullberg 1994, 1996), is very similar in the four modes of comparisons, supporting the versatility of this gene in phylogenetic analyses.

Table 6 summarizes the molecular differences among the four species-pairs studied. The order among the species-pairs is the same irrespective of the comparison applied, although the magnitude of the differences may show some variation. Thus the aa difference between the donkey and the horse (1.9%), relative to that of the seals (1.6%), is limited compared with both total nt difference (6.9%/3.5%) and conservative nt difference (1.2%/0.7%). With respect to ratios for total nt substitution according to codon position, the value for codon position 3 (25.7) between the donkey and the horse is considerably higher than that of the other three comparisons. Also with respect to ratios for conservative nt differences the third codon position values for donkey/horse are higher than those of the other three comparisons. We believe, however, that these differences represent normal variation among recent mammalian divergences because there is no contrasting pattern between the most recent divergence (the seals) and the three more distant ones. The transition/transversion (Ti/Tv) ratios in third codon position show some variation. This variation does not appear to be directly related to molecular differences, suggesting that transitional saturation is limited or nonexistent among the four comparisons.

The Dating of the Evolutionary Separation of the Donkey and the Horse

The availability of the mtDNA molecules of the donkey and the horse makes it possible to examine the donkey/horse divergence on the basis of complete mtDNA data. Comparisons with external standards (Krettek et al. 1995) show that the mtDNAs of equiids (horse) evolve slightly slower than those of seals and whales. Applying an external reference, the artiodactyl/cetacean divergence set at 60 MYA (Arnason and Gullberg 1996), the divergence between the two seals has been dated at ≈7 MYA and that between the two whales at ≈11 MYA (Arnason et al. 1996b). The difference between the horse and the donkey relative to these two species-pairs varies, however, with respect to the mode of comparison (total

Table 4. Extended

Gene	Amino acid difference—total/chemical properties			
	E.as/E.ca	*P.vi/H.gr*	*B.ph/B.mu*	*Homo/P.tr*
COI	(1) 0.2/0.2 (3)	(6) 1.0/0.8 (5)	(2) 0.6/0.2 (3)	(2) 1.2/0.8 (1)
COIII	(2) 0.4/0.0 (1)	(3) 0.8/0.0 (2)	(3) 0.8/0.0 (2)	(4) 2.7/1.5 (3)
COII	(3) 0.4/0.4 (5)	(2) 0.4/0.4 (3)	(5) 1.3/1.3 (5)	(3) 2.2/1.3 (2)
NADH3	(5) 0.9/0.0 (2)	(5) 0.9/0.9 (6)	(1) 0.0/0.0 (1)	(11) 6.1/5.2 (11)
NADH1	(7) 1.6/0.9 (7)	(4) 0.9/0.6 (4)	(4) 1.3/0.9 (4)	(9) 5.0/3.8 (8)
NADH4L	(8) 2.0/1.0 (8)	(1) 0.0/0.0 (1)	(11) 7.1/3.1 (10)	(1) 1.0/2.0 (4)
NADH4	(6) 1.5/0.9 (6)	(7) 1.7/1.1 (7)	(6) 3.3/2.0 (7)	(7) 4.7/2.6 (6)
Cytb	(4) 0.8/0.3 (4)	(10) 2.4/1.1 (8)	(7) 3.4/1.3 (6)	(13) 7.1/5.5 (13)
NADH2	(11) 3.7/2.6 (11)	(9) 2.0/1.2 (9)	(9) 4.9/3.2 (11)	(5) 4.0/3.5 (7)
NADH5	(9) 2.8/1.8 (9)	(8) 1.8/1.3 (10)	(8) 3.5/2.0 (8)	(12) 6.6/5.5 (12)
ATPase6	(10) 3.1/1.8 (10)	(11) 2.7/2.2 (12)	(10) 5.3/4.0 (12)	(84) 4.9/4.0 (9)
NADH6	(12) 5.7/2.9 (12)	(12) 2.9/1.7 (11)	(12) 7.3/2.3 (9)	(6) 4.6/2.3 (5)
ATPase8	(13) 8.8/7.5 (13)	(13) 4.5/3.0 (13)	(13) 7.9/7.9 (13)	(10) 5.8/4.4 (10)
Mean	1.9/1.4	1.6/1.2	3.0/2.0	4.4/3.3

Table 5. Mean ranking values and positions of individual genes according to increasing difference in four pairwise comparisons—donkey and horse, harbor and grey seals, fin and blue whales, *Homo* and common chimpanzee[a]

Gene	Mean position of nucleotide difference		Mean position of amino acid difference	
	Total	Conservative	Total	Chemical
COI	4.3 (2/3)	2.3 (1)	2.8 (1)	3.0 (2)
COIII	5.5 (5)	4.0 (3)	3.0 (2)	2.0 (1)
COII	6.3 (7)	3.0 (2)	3.3 (3)	3.8 (3)
NADH3	6.0 (6)	5.0 (4/5)	5.5 (5)	5.0 (4)
NADH1	3.8 (1)	5.0 (4/5)	6.0 (6)	5.8 (5/6)
NADH41	4.8 (4)	6.0 (6)	5.3 (4)	5.8 (5/6)
NADH4	11.0 (13)	6.8 (7)	6.5 (7)	6.5 (7)
Cytb	8.5 (9/10)	9.0 (9)	8.5 (8/9)	7.8 (8)
NADH2	10.5 (11/12)	8.3 (8)	8.5 (8/9)	9.5 (10)
NADH5	10.5 (11/12)	9.5 (11)	9.3 (10)	9.8 (11)
ATPase6	7.3 (8)	9.3 (10)	9.8 (11)	10.8 (12)
NADH6	8.5 (9/10)	10.8 (12)	10.5 (12)	9.3 (9)
ATPase8	4.3 (2/3)	12.3 (13)	12.3 (13)	12.3 (13)

[a] Position values show the mean of ranking position for each gene among the comparisons of the four pairs of species in Table 4. Figures in parenthesis show the ranking position within each column.

nt difference, conservative nt difference, aa difference). We see, therefore, no reason to attempt to date the horse/donkey divergence any narrower than 8–10 MYA. With respect to the rRNA data in Table 6, the 12S rRNA difference yield datings that are in line with those for total nt difference in peptide coding genes, whereas the 16S rRNA differences conform with those for conservative nt substitutions.

The sequence of the complete cytochrome *b* gene of Grévy's zebra, *E. grevyi*, has been reported (Irwin et al. 1991). The total nt difference between the cyt *b* genes of the donkey and the horse is 7.7%. The percent conservative nt difference is 1.1%, and aa difference 0.8%. The corresponding values for the zebra and the horse are 8.2%, 1.0%, and 1.1%, and those for the zebra and the donkey are 5.9%, 0.8%, and 1.3%. The different values do not provide an unequivocal picture of the relationship among the three species, but it is likely that in close relationships like these the stochastic error in comparisons based on total nt difference is less than that in the other two modes of comparison. The total nt difference suggests, consistent with George and Ryder (1986), that there is a closer relationship between the donkey and the zebra than there is between either of these species and the horse.

Discussion

Besides providing details of the molecular relationship between the mtDNAs of the donkey and the horse, the present study examined the rate of molecular evolution among peptide coding genes of four recent divergencies (harbor seal/grey seal, horse/donkey, fin whale/blue whale, *Homo*/common chimpanzee). Four modes of comparison were performed within each species-pair and the ranking position for each gene was determined. Calculation of Kendall's coefficient of rank correlation showed that the ranking position of individual genes was somewhat irregular among the four species-pairs. As expected this irregularity was most strongly pronounced in comparisons based on total nt substitution, but also apparent to some extent in the three other, and more conservative, modes of comparison. The analyses showed that the evolutionary rate of the same peptide-coding gene, relative to other mtDNA peptide-coding genes, may differ considerably among recent divergences of different lineages such as equiids, seals, balaenopterid whales, and hominoids. The results emphasize that findings based on limited sets of sequence data for resolving and dating evolutionary divergences among, or within,

Table 6. Comparison of mtDNA in four closely related mammalian species-pairs: donkey/horse, harbor/grey seals, fin/blue whales, and human/chimpanzee[a]

	Seals	Donkey/horse	Whales	Human/chimpanzee
Percent nt different	3.5	6.9	7.5	8.5
Percent cons. nt difference	0.7	1.2	1.5	2.1
Percent aa difference	1.6	1.9	3.0	4.4
Ratio total nt substitution	2.7:1.0:16.0	3.5:1.0:25.7	4.2:1.0:17.0	2.8:1.0:12.3
Ratio cons. nt substitution	1.6:1.0:0.9	1.5:1.0:2.1	1.6:1.0:1.5	1.6:1.0:0.8
Ratio 3rd condon pos. Ti/Tv	16.6	11.2	10.4	13.7
Percent nt difference of 12S rRNA	2.2/1.9/0.2/0.1	4.9/3.9/0.7/0.3	4.9/3.7/0.7/0.5	4.5/3.9/0.3/0.3
Percent nt difference of 16S rRNA	2.6/1.8/0.4/0.4	3.7/2.9/0.6/0.2	5.3/4.3/0.8/0.2	5.5/4.8/0.6/0.1

[a] Percent nt (nucleotide) difference does not include control regions. Percent cons. (conservative) nt difference is limited to peptide-coding genes. Percent aa (amino acid) difference is based on concatenated protein sequence of all peptide-coding genes. Ratios for total and conservative nt substitution according to codon position are based on all 13 peptide-coding genes; 2nd codon position has arbitrarily been given the value 1.0. Percent rRNAs differences are arranged as: total/Ti/Tv/gaps

distantly related lineages should be taken with caution unless supported by extensive species sampling. The present analyses of close evolutionary relationships are consistent with, and complementary to, earlier findings of distant eutherian relationships (Arnason and Johnsson 1992; Cao et al. 1994) which showed that different mtDNA genes provide different topologies for the relationship among Rodentia (mouse, rat), Primates (*Homo*), Carnivora (harbor seal), Artiodactyla (cow), and Cetacea (fin whale).

Consultation of Table 6 shows that the molecular difference is least between the seals, followed by horse/donkey, the two whales, and *Homo*/common chimpanzee. This order is not expressed, however, in the 12S rRNA gene, where horse/donkey exhibit the same difference as the two whales, and where the difference between *Homo*/common chimpanzee is less than that between these two species-pairs. Table 6 also gives an account of the rates of nt substitution, both total and conservative, according to codon position in four species-pairs belonging to four different eutherian orders. The similarities between the data of the seals (least molecular difference) and those of *Homo*/chimpanzee (largest molecular difference) suggest that none of the four pairwise comparisons is affected by mutational saturation.

The proposed dating of the divergence between donkey and horse, 8–10 MYA, is distinctly earlier than both paleontological data (Simpson 1951; Lindsay et al. 1980) and the datings proposed by restriction endonuclease analysis of the equiids (George and Ryder 1986). It has been demonstrated recently (Ohland et al. 1995) by simulated restriction analyses of complete mtDNA sequences (where actual sequence differences are known), that real (experimentally performed) restriction analyses may provide pictures of sequence divergence that deviate quite considerably from true divergences. The data on sequence differences obtained by George and Ryder (1986) are, nevertheless, in line with the present findings. Our dating of the evolutionary separation between the horse and donkey is, however, considerably earlier than proposed by them. In their calculations George and Ryder (1986) applied a divergence rate of 2% per million years as proposed by Brown et al. (1979). It is probable that this value substantially overestimates the rate of evolutionary divergence among equiids.

Acknowledgments. We express our thanks to Dr. Stuart W. J. Reid, University of Glasgow Veterinary School, UK, for kindly providing us with samples, also to Prof. Einar Arnason for valuable comments on the manuscript, and to Prof. Jan Lanke for advice on the statistical analyses. The work was supported by grants from the Swedish Natural Sciences Research Council, the Erik-Philip-Sörensen Foundation, and the Swedish Institute.

References

Anderson S, Bankier AT, Barrell BG, deBruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Ros BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

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

Arnason U, Gullberg A (1993) Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. J Mol Evol 37:312–322

Arnason U, Gullberg A (1994) Relationship of baleen whales established by cytochrome *b* gene sequence comparison. Nature 367: 726–728

Arnason U, Gullberg A (1996) Cytochrome b nucleotide sequences and the identification of five primary lineages of extant cetaceans. Mol Biol Evol 13:407–417

Arnason U, Johnsson E (1992) The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina.* J Mol Evol 34:493–505

Arnason U, Gullberg A, Widegren B (1991a) The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus.* J Mol Evol 33:556–568

Arnason U, Spilliaert R, Palsdottir A, Arnason A (1991b) Molecular hybridization of hybrids between the two largest whale species, the blue (*Balaenoptera musculus*) and the fin whale (*B. physalus*). Hereditas 115:183–189

Arnason U, Gullberg A, Johnsson E, Ledje C (1993) The nucleotide

sequence of the mitochondrial DNA molecule of the grey seal, *Halichoerus grypus*, and a comparison with mitochondrial sequences of other true seals. J Mol Evol 37:323–330

Arnason U, Bodin K, Gullberg A, Ledje C, Mouchaty S (1995) A molecular view of pinniped relationships with particular emphasis on the true seals. J Mol Evol 40:78–85

Arnason U, Xu X, Gullberg A (1996a) Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on non-chimaeric sequences. J Mol Evol 42: 145–152

Arnason U, Gullberg A, Janke A, Xu X (1996b) Pattern and timing of evolutionary divergences among hominoids based on analyses of complete mtDNAs. J Mol Evol (in press)

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180

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

Cao Y, Adachi J, Janke A, Pääbo S, Hasegawa M (1994) Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of tree based on a single gene. J Mol Evol 39:519–527

Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisa E, Saccone C (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 28:497–516

George MJ, Ryder OA (1986) Mitochondrial DNA evolution in the genus *Equus.* Mol Biol Evol 3(6):535–546

Gribskov M, Burgess RR (1986) Sigma factors from E. coli, B. subtilis, phage SPO1, and phage T4 are homologous proteins. Nucleic Acids Res 14(16):6745–6763

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92: 532–536

Irwin DM, Arnason U (1994) Cytochrome b gene of marine mammals: phylogeny and evolution. J Mamm Evol 2:37–55

Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome *b* gene of mammals. J Mol Evol 32:128–144

Janke A, Feldmaier-Fuchs G, Thomas WK, Haeseler A, Pääbo S (1994) The marsupial mitochondrial genome and the evolution of placental mammals. Genetics 137:243–256

Janke A, Gemmell NJ, Feldmaier-Fuchs G, Haesler A, Pääbo S (1996) The mitochondrial genome of a monotreme—the platypus (*Ornitorhynchus anatinus*). J Mol Evol 42:153–159

Krettek A, Gullberg A, Arnason U (1995) Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphla. J Mol Evol 41:952–957

Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J Mol Evol 37:380–398

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

Ma DP, Zharkikh A, Graur D, VandeBerg JL, Li WH (1993) Structure and evolution of opossum, guinea pig and porcupine cytochrome *b* genes. J Mol Evol 36:327–334

Ohland DP, Harley EH, Best PB (1995) Systematics of cetaceans using restriction site mapping of mitochondrial DNA. Mol Phylogenet Evol 4:10–19

Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474

Ozawa T, Tanaka M, Sugiyama S, Ino H, Ohno K, Hattori K, Ohbayashi T, Ito T, Deguchi H, Kawamura K, Nakase Y, Hashiba K (1991) Patients with idiopathic cardiomyopathy belong to the same mitochondrial DNA gene family of Parkinson's disease and mitochondrial encephalomyopathy. Biochem Biophys Res Commun 177: 518–525

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

Spilliaert R, Vikingsson G, Arnason U, Palsdottir A, Sigurjonsson J, Arnason A (1991) Species hybridization between a female blue whale (*Balaenoptera musculus*) and a male fin whale (*B. physalus*): molecular and morphological documentation. J Hered 82:269–274

Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. Int Rev Cytol 141:173–216

Xu X, Arnason U (1994) The complete mitochondrial DNA sequence of the horse, *Equus caballus:* extensive heteroplasmy of the control region. Gene 148:357–362

Xu X, Arnason U (1996) A complete sequence of the mitochondrial genome of the western lowland gorilla. Mol Biol Evol 13:691–698