The Complete Mitochondrial Sequence of *Tarsius bancanus*: Evidence for an Extensive Nucleotide Compositional Plasticity of Primate Mitochondrial DNA

Jürgen Schmitz, Martina Ohme, and Hans Zischler
Primate Genetics, German Primate Center

Inconsistencies between phylogenetic interpretations obtained from independent sources of molecular data occasionally hamper the recovery of the true evolutionary history of certain taxa. One prominent example concerns the primate infraordinal relationships. Phylogenetic analyses based on nuclear DNA sequences traditionally represent Tarsius as a sister group to anthropoids. In contrast, mitochondrial DNA (mtDNA) data only marginally support this affiliation or even exclude Tarsius from primates. Two possible scenarios might cause this conflict: a period of adaptive molecular evolution or a shift in the nucleotide composition of higher primate mtDNAs through directional mutation pressure.

To test these options, the entire mt genome of *Tarsius bancanus* was sequenced and compared with mtDNA of representatives of all major primate groups and mammals. Phylogenetic reconstructions at both the amino acid (AA) and DNA level of the protein-coding genes led to faulty tree topologies depending on the algorithms used for reconstruction.

We propose that these artifactual affiliations rather reflect the nucleotide compositional similarity than phylogenetic relatedness and favor the directional mutation pressure hypothesis because: (1) the overall nucleotide composition changes dramatically on the lineage leading to higher primates at both silent and nonsilent sites, and (2) a highly significant correlation exists between codon usage and the nucleotide composition at the third, silent codon position. Comparisons of mt genes with mt pseudogenes that presumably transferred to the nucleus before the directional mutation pressure took place indicate that the ancestral DNA composition is retained in the relatively fossilized mtDNA-like sequences, and that the directed acceleration of the substitution rate in higher primates is restricted to mtDNA.

Introduction

The phylogenetic relationship of tarsiers to other primates has been a source of debate for many decades. Tarsius is the only remaining genus of a formerly diverse group of Eocene tarsiiformes that shares characteristics of both prosimians and anthropoids. These features complicate the reconstruction of the evolutionary history (Fleagle 1999, pp. 120–122).

A general conflict exists at the molecular level between nuclear and mitochondrial DNA (mtDNA) data. Whereas nuclear DNA tends to cluster tarsiers together with anthropoids (Goodman et al. 1998), the mt cytochrome *c* oxidase subunit II reveals only a weak affinity of tarsiers to anthropoids (49.6% of bootstrap replications [BR] [Adkins and Honeycutt 1994]). On the basis of cytochrome *b* sequences, tarsiers are even placed apart from the remaining primates (Andrews, Jermiin, and Easteal 1998). However, in an extensive screening of orthologous retroposons as molecular cladistic markers, the sister taxon relationship of Tarsius to anthropoids was settled (Schmitz, Ohme, and Zischler 2001).

To date, it remains open as to which of the evolutionary forces caused the tarsiers to shift from their historical place in the mtDNA tree. Adkins, Honeycutt, and Disotell (1996) reported a rapid burst of AA replacements in COII on the lineage leading to higher primates after Tarsius diverged. Most of the changes occurred independently along the New World monkey lineage and the branch leading to the Old World monkeys. The authors indicate that the rate acceleration is mainly restricted to nonsynonymous sites; and moreover, it is responsible for structural changes of cytochrome *c* oxidase in higher primates. Similar results were obtained by comparing substitution rates of nonsynonymous changes versus synonymous transversions at fourfold degenerate sites of cytochrome *b* (Andrews, Jermiin, and Easteal 1998). From the relative rate test the authors concluded that anthropoid primate cytochrome *b* has undergone an episode of adaptive evolution. The lack of significant changes of fourfold degenerate transversions was taken as an argument against a more general mutation-based rate acceleration in the lineage leading to higher primates. Furthermore, the authors propose an episode of concerted evolution of cytochrome *b* and COII in higher primates. Positive selection at the AA level is one of the potential forces challenging variation at the protein level. A more general phenomenon is the alteration of the nucleotide composition under the influence of directional mutation pressure (Sueoka 1992). Variation in nucleotide composition is preferentially pronounced at the synonymous codon positions with little effect on the AA content of the encoded protein. If both the synonymous and nonsynonymous codon positions are affected by compositional bias at the DNA level, then proteins will change their AA composition in the direction predicted by the underlying nucleotide bias and selective constraints (Singer and Hickey 2000).

In order to obtain a more conclusive picture of the forces responsible for the peculiar evolution of primate mtDNA, we compared the entire mt genome of *Tarsius bancanus* with that of other primates and nonprimate mammals. This data set now comprises representatives of all major primate groups, including strepsirrhini, tarsiiformes, platyrrhini, and catarrhini.

Materials and Methods
Isolation and Amplification of mtDNA

A tissue sample of the Western tarsier (*T. bancanus*) was provided by Y. Rumpler, Les Hôpitaux Universitaires de Strasbourg, France.

Because an isolation of enriched mtDNA was precluded by the limited amount and quality of the available tissue, we extracted total DNA from 50 mg using the QIAamp Mini Kit (QIAGEN). To avoid an inadvertent PCR amplification of possible mt pseudogenes, we chose a strategy in which long-range PCRs of five approximately 3–4 kbp overlapping fragments, exceeding the average length of nuclear integrations of mtDNA described to date (Zhang and Hewitt 1996), were performed initially. The resulting PCR fragments and the total DNA were used as templates in a second round of PCRs, generating 19 overlapping fragments with a size of about 1 kbp. Each fragment was PCR-amplified at least twice and then processed independently. Both the long PCR fragments and the 19 one-kbp fragments were ligated into pGEM-T Vectors (PROMEGA) and electroporated into TOP10 electrocompetent cells (INVITROGEN). The sequences of both insert strands were determined using a LI-COR 4200 automated DNA sequencer. The complete mt genome of *T. bancanus* was deposited in GenBank under the accession number AF348159.

For pseudogene comparisons, we identified nuclear integrations of mtDNA in humans by querying the database with a human cytochrome *b* sequence. Phylogenetic reconstructions based on comparisons with mt sequences of Tarsius and published anthropoid representatives suggested that one of these integrations took place before the divergence of anthropoids. On the basis of the 5′ and 3′ nuclear sequences that flank the mtDNA-like fragment, two PCR primers were constructed. These primers (numt-1: 5′ AGTCGGTAATACTAGATTTAT GACAGT 3′; numt-2: 5′ TCAATTAATCACAGAACC AGCAT 3′) were used in a standard PCR to amplify and sequence the homologous fragment in *Macaca mulatta*. Cloning and sequencing was done as outlined previously (GenBank accession number AF378365).

Incorporated GenBank Entries

Additional mt genomes were obtained from published sources: human, X93334 (Arnason, Xu, and Gullberg 1996); common chimpanzee, D38113; gorilla, D38114; orangutan, D38115 (Horai et al. 1995); gibbon, X99256 (Arnason, Gullberg, and Xu 1996); hamadryas baboon, Y18001 (Arnason, Gullberg, and Janke 1998); barbary macaque, AJ309865; white-fronted capuchin, AJ309866; slow loris, AJ309867; Tupaia, AF217811 (Schmitz, Ohme, and Zischler 2000), armadillo, Y11832

(Arnason, Gullberg, and Janke 1997); bat, AF061340 (Pumo et al. 1998); megabat, NC_002612 (Nikaido et al. 2000); pig, AJ002189 (Ursing and Arnason 1998); cow, J01394 (Anderson et al. 1982); fin whale, X61145 (Arnason, Gullberg, and Widegren 1991); horse, X79547 (Xu and Arnason 1994); cat, U20753 (Lopez, Cevario, and O'Brien 1996); dog, U96639 (Kim et al. 1998); gray seal, X72004 (Arnason et al. 1993); harbor seal, X63726 (Arnason and Johnsson 1992); mouse, J01420 (Bibb et al. 1981); rat, X14848 (Gadaleta et al. 1989); opossum, Z29573 (Janke et al. 1994); platypus, X83427 (Janke et al. 1996). For comparison with mt pseudogenes, we retrieved full-length cytochrome *b*-like sequences located in the nucleus from human accessions AC002087 and NT_006654 and the cytochrome *b* sequence of *Lemur catta* (U38271).

Data Analysis

Protein-coding and rDNA sequences were aligned by Clustal X (Thompson et al. 1997). For protein-coding genes we used the alignment of Schmitz, Ohme, and Zischler (2000) as a profile for adding the *T. bancanus* sequence. rDNA alignments were adjusted to secondary structure models provided by Hixon and Brown (1986) (12S rDNA) and Gutell and Fox (1988) (16S rDNA), respectively.

Phylogenetic reconstructions for both DNA and AA sequences were performed by three different approaches: (1) maximum parsimony (MP, heuristic search, 1,000 BR), (2) distance-based methods (neighbor-joining [NJ], 1,000 BR) as implemented in PAUP* version 4.0b8 (Swofford 2000) and PHYLIP version 3.573c (Felsenstein 1995), and (3) maximum likelihood (ML) with the discrete gamma-distribution model of rate heterogeneity over sites (1,000 puzzling steps [QPS]) included in PUZZLE version 5.0 (Strimmer and von Haeseler 1996). Secondary structures of the sequenced tRNAs were confirmed using the compilation of Sprinzl et al. (1998). RRTree version 1.1 was used to compare lineage-specific substitution rates (Robinson et al. 1998).

To split the rDNA sequences in conserved and variable regions, we used Gblocks version 0.73b (Castresana 2000) applying the default settings for rDNA. MFOLD (Zuker, Mathews, and Turner 1999) was used to identify and verify secondary structure formation. Reconstruction of ancestral character states was performed by MP (Swofford 2000). A multivariate analysis of codon usage was carried out with codonW version 1.4.2 (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html).

Results
General Features of the mt Genome of *T. bancanus*

The length of the complete mt genome of *T. bancanus* is 16,927 nt. This value is not absolute because identical, tandemly arranged repeats of 22 nt (TACA-CCCATGCGTACACGCACG) occur in variable numbers (*n* = 4–15 in different PCR products) between the conserved sequence blocks CSB1 and CSB2 of the con-

Table 1
Organization of the *Tarsius bancanus* Mitochondrial Genome

Gene[a]	From	To	Size	Start Codon	Stop Codon[b]	Spacer-Overlap
tRNA-Phe	1	67	67			
12S rRNA	68	1025	958			
tRNA-Val	1026	1090	65			
16S rRNA	1091	2660	1,570			
tRNA-Leu (UUR)	2661	2735	75			2-base spacer
NADH 1	2738	3694	957	ATG	TAA	6-base spacer
tRNA-Ile	3701	3769	69			3-base overlap
tRNA-Gln (L)	3767	3839	73			1-base spacer
tRNA-Met	3841	3909	69			
NADH 2	3910	4953	1,044	ATT	Tag	2-base overlap
tRNA-Trp	4952	5019	68			10-base spacer
tRNA-Ala (L)	5030	5099	70			3-base spacer
tRNA-Asn (L)	5103	5175	73			
Origin-L (L)	5176	5208	33			
tRNA-Cys (L)	5209	5274	66			
tRNA-Tyr (L)	5275	5340	66			7-base spacer
COI	5348	6889	1,542	ATG	TAA	7-base spacer
tRNA-Ser (UCN) (L)	6897	6965	69			7-base spacer
tRNA-Asp	6973	7041	69			
COII	7042	7725	684	ATG	TAA	4-base spacer
tRNA-Lys	7730	7797	68			
ATPase8	7798	8001	204	ATG	TAA	43-base overlap
ATPase6	7959	8639	681	ATG	TAa	1-base overlap
COIII	8639	9422	784	ATG	T–	1-base overlap
tRNA-Gly	9422	9493	72			1-base overlap
NADH 3	9493	9839	347	ATA	TA-	
tRNA-Arg	9840	9907	68			1-base spacer
NADH 4L	9909	10205	297	ATG	TAA	7-base overlap
NADH 4	10199	11576	1,378	ATG	T–	
tRNA-His	11577	11645	69			
tRNA-Ser (AGY)	11646	11705	60			1-base overlap
tRNA-Leu (CUN)	11705	11774	70			
NADH 5	11775	13586	1,812	ATA	TAA	4-base overlap
NADH 6 (L)	13583	14095	513	ATG	TAG	1-base spacer
tRNA-Glu (L)	14097	14165	66			3-base spacer
Cytb	14169	15308	1,140	ATG	TAA	
tRNA-Thr	15309	15377	69			5-base spacer
tRNA-Pro (L)	15383	15448	66			
Control region	15449	16927	1,479			

[a] (L) indicates a gene encoded on the L-strand.
[b] Lowercase letters indicate possible overlap with the beginning of the following gene. The 22 tRNA sequences do not vary substantially from the typical mammalian secondary structures as compiled by Sprinzl et al. (1998).

trol region. Heteroplasmy of this region is a common phenomenon in eutherians (Sbisa et al. 1997).

The entire L-strand base composition is: A, 33.0%; C, 26.6%; G, 12.4%; and T, 28.0%. The gene order matches the organization in other eutherian mt genomes. The information about gene localization, initiation and termination, overlap, and spacer sequences is shown in table 1. The length of the concatenated rRNA genes is 2,663 nt. By comparing with other eutherians we determined 2,257 nt in conserved and 406 nt in poorly alignable regions.

Phylogenetic Evaluation

In order to characterize the misleading signals that give rise to the conflicting phylogenetic position of *T. bancanus* with respect to other primates, we included mt sequences of all major primate groups represented by the catarrhines (human, common chimpanzee, gorilla, orangutan, gibbon, baboon, and barbary macaque),

the platyrrhines (white-fronted capuchin monkey), and the strepsirrhines (slow loris); as nonprimate representatives we analyzed tree shrew and armadillo, the cetartiodactyla pig, cow, whale; the horse; the carnivora cat, dog, gray seal, and harbor seal; and the rodentia mouse and rat. For phylogenetic reconstructions we concatenated all H-strand–encoded AA and DNA sequences, respectively. The opossum and platypus mt genome was used to root the phylogenetic trees.

On the basis of AA replacements, only the MP method of reconstruction yielded the correct topology, with Tarsius and the anthropoids forming a monophyletic group (55% BR; fig. 1*A*). ML with eight rate categories and the mtREV24 model of substitution gave rise to a topology grouping Tarsius and the slow loris together (95% of QPS; fig. 1*B*). NJ using the Dayhoff PAM matrix retrieves the same tree with 68% BR for a Tarsius-slow loris clade.

Concerning the analyses at the DNA level, which were restricted to the first and second codon positions,
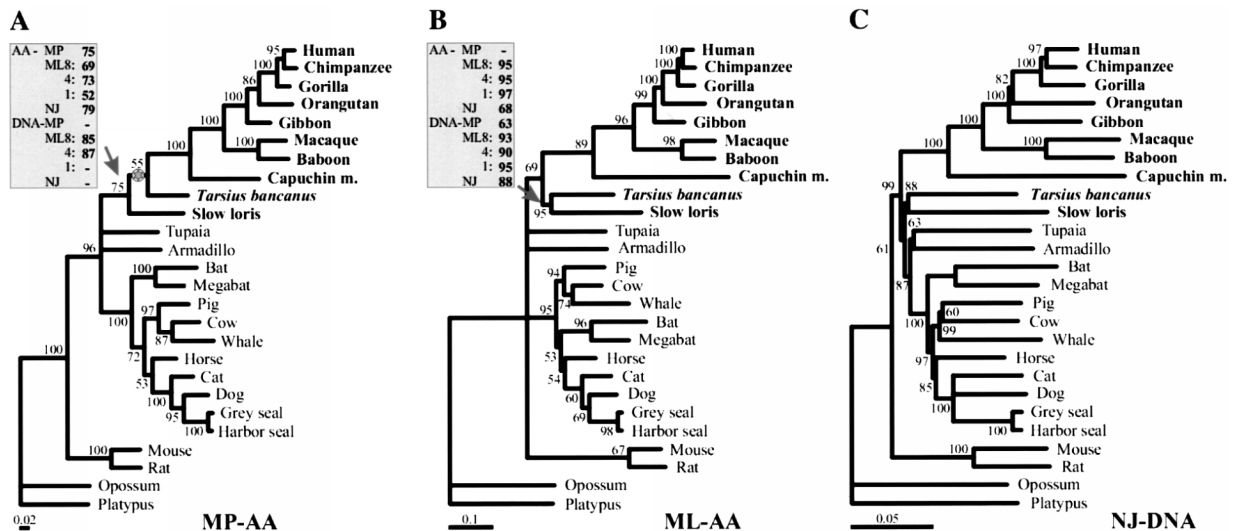
Fig. 1.—Phylogenetic tree reconstruction based on: (A) amino acid sequences of the 12 H-strand genes (maximum likelihood reconstruction with eight rate categories); (B) ML reconstruction; (C) first and second codon positions of the 12 H-strand genes (NJ reconstruction). Branch lengths represent amino acid or nucleotide substitutions per site. Percentage BR or QPS values are indicated at the corresponding nodes. Primates are shown in bold face. The opossum and platypus were used as an outgroup. The asterisk labels a branch that is consolidated by three SINE markers (Schmitz, Ohme, and Zischler 2001).

ML, MP, and NJ tree reconstructions support a cluster of Tarsius and slow loris with 93% QPS, 63% and 88% BR, respectively (fig. 1). The NJ reconstruction displays the Tarsius-slow loris clade inside the nonprimate mammals (61% BR; fig. 1C).

In addition, applying various distance-based corrections, including LogDet which takes base compositional differences between taxa into account, the conflicting tree topology shown in figure 1C could not be resolved. We assume that the rate heterogeneity across sites represents a problem in LogDet-based distance estimations (see also Waddell et al. 1999).

Concerning both AA- and DNA-based phylogenetic reconstructions, none of the estimated conflicting topologies shown in figure 1B and C were significantly different from the canonical tree, with Tarsius representing the sister group to the Anthropoidea and the slow loris branching off first among primates (fig. 1A) at a 5% level of significance (Kishino and Hasegawa 1989).

Relative Rate Test

To check for lineage-specific significant differences in the evolutionary rate, relative rate tests of nonsynonymous and synonymous substitutions were carried out on the 12 concatenated protein-coding DNAs (table 2). As recommended for the relative rate test, we restricted our analysis to those species whose phylogenetic relationships are generally accepted. We grouped the species in the Anthropoidea, Tarsius, slow loris, tree shrew, cetferungulates, and opossum.

As an outgroup, we used the respective closest sister taxon to the compared taxa. However, although the relative rate test detected a highly significant change in the nonsynonymous substitution rate along the lineage leading to the Anthropoidea, the rate of synonymous changes remained homogeneous. Tarsius, in comparison with cetferungulates, did not exhibit a significant rate ratio. Concerning the synonymous substitution rate among the compared taxa, a significant change was only observed between human and common chimpanzee.

Table 2
Relative Rates of Synonymous (Ks) and Nonsynonymous (Ka) Substitutions

Taxon 1	Taxon 2	Outgroup	P Ks	Included Sites	P Ka	Included Sites
Human	Chimpanzee	Gorilla	*	TI + TV/3991		TI + TV/6854
Anthropoidea . . .	Tarsius	Tupaia		TI + TV/3697	**	TI + TV/7112
Anthropoidea . . .	Tarsius	Tupaia		TV/1916[a]	**	TV/7112[b]
Anthropoidea . . .	Slow loris	Tupaia		TI + TV/3709	**	TI + TV/7100
Tarsius	Slow loris	Tupaia		TI + TV/3730		TI + TV/7079
Anthropoidea . . .	Cetferungulata	Opossum		TI + TV/3610	**	TI + TV/7166
Anthropoidea . . .	Cetferungulata	Opossum		TV/1867[a]	**	TV/7166[b]
Tarsius	Cetferungulata	Opossum		TI + TV/3610		TI + TV/7166
Tarsius	Cetferungulata	Opossum		TV/1848[a]		TV/7166[b]
Slow loris	Cetferungulata	Opossum		TI + TV/3658		TI + TV/7118

NOTE.—Ks: number of synonymous substitutions per synonymous site including transitions and transversions (TI + TV). Ka: number of nonsynonymous substitutions per nonsynonymous site including transitions and transversions (TI + TV). Significance level: $P = 0.05$–0.01 (*); $P < 10^{-6}$ (**).

[a] Number of synonymous transversions per fourfold degenerate site (TV).

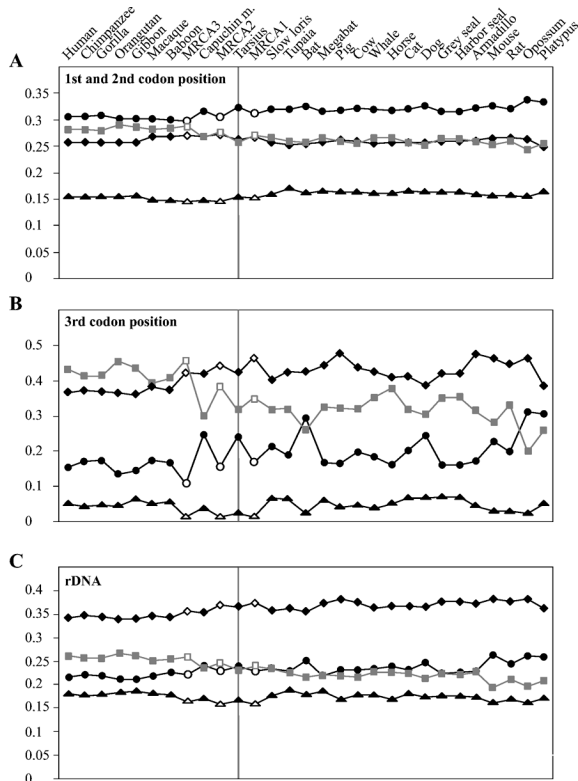[b] Number of nonsynonymous transversions per nonsynonymous site (TV).

FIG. 2.—Nucleotide frequencies of 26 mammalian mtDNAs and three estimated MRCA sequences. (*A*) Nucleotide frequencies for the first and second codon position of the 12 H-strand genes. (*B*) Third codon position of the 12 H-strand genes. (*C*) Concatenated rRNA genes. The adenosine base frequency is represented by rhombs, cytosines by gray rectangles, guanines by triangles, and thymines by circles. Open symbols represent the respective frequencies in the MRCAs of Tarsius and the anthropoids (MRCA1), the anthropoids (MRCA2), and the catarrhines (MRCA3).

Base Composition

A comparison of the respective base compositions obtained from 26 mammalian species is given in figure 2. To obtain a more precise description of the course of compositional changes during primate evolution, we computed the ancestral sequences of the most recent common ancestors of the anthropoids and Tarsius (MRCA1), anthropoids (MRCA2) and catarrhines (MRCA3), according to parsimony criteria. Separate estimations have been done for the first and second codon positions of the 12 concatenated H-strand protein-coding DNAs, the third codon position, and the concatenated rRNA genes (fig. 2). The nucleotide composition among and between nonprimate mammals, Tarsius, and slow loris is similar; however, it changes dramatically along the lineage leading to higher primates. Whereas the gradual shift of the nucleotide composition along the lineage leading to the catarrhines is verified in the situation found in the MRCA1 and MRCA2, the capuchin monkey reverts, as an autapomorphy, its composition toward the nonprimate situation. This alteration is most clearly exemplified by a compositional shift from T to C and A to C.
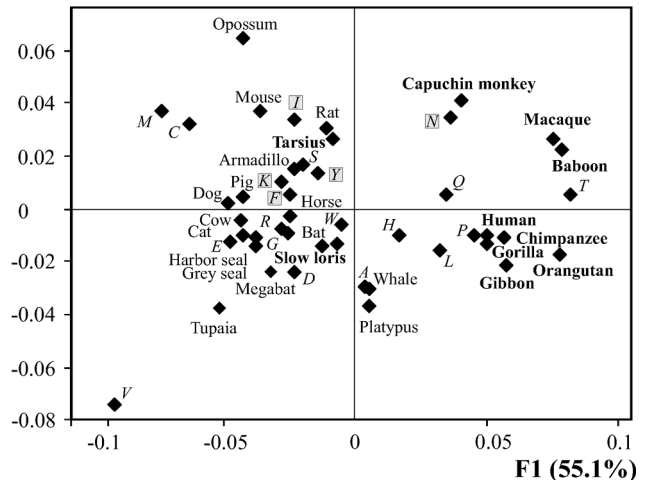

FIG. 3.—Correspondence analysis of the mt 12 H-strand protein-coding regions from 26 mammals as a function of amino acid composition. Note the separated position of Tarsius compared with other primates and the shift in AA preference. F1 and F2, first and second factorial axes represent 55.1% and 18.1% of the total variability in the correspondence analysis, respectively. A+T-rich AAs are highlighted in gray boxes.

AA Usage

We performed a correspondence analysis of AA usage to determine whether the observed compositional changes of nucleotides are associated with the AA content of the primate mt polypeptides. From figure 3 it is obvious that Tarsius and slow loris cluster apart from the remaining primates closest to nonprimate mammals. Thus, in contrast to higher primates, Tarsius and slow loris exhibit an affinity toward the AAs isoleucine (I), lysine (K), phenylalanine (F), tyrosine (Y), which are encoded by A+T-rich codons (Foster, Jermiin, and Hickey 1997). Again, the capuchin monkey cannot be unequivocally assigned to either the typical anthropoid or nonprimate pattern.

Regression Analysis

To determine whether the compositional changes of the nucleotide and AA content in higher primates are caused by directional mutation pressure or a result of positive selection, we performed a correlation analysis. If the nucleotide bias affects both the synonymous and nonsynonymous sites in protein-coding genes, a phenomenon that could not solely be explained by positive selection, then we would expect the proteins to change their AA composition in line with the underlying nucleotide bias (see also Singer and Hickey [2000]). Figure 4*A* indicates a highly significant correlation between the proportion of A+T nucleotides (all codon positions included), which is mostly pronounced in the catarrhines and Tarsius, and the frequency of A+T-rich AAs (phenylalanine [F], tyrosine [Y], methionine [M], isoleucine [I], asparagine [N], and lysine [K]; $r^2 = 0.59$, $P < 10^{-4}$). The correlation between A+T frequency at third codon positions and the usage of A+T-rich AAs is also highly significant ($r^2 = 0.52$, $P < 10^{-4}$; see fig. 4*B*).

A

Proportion of FYMINK amino acids
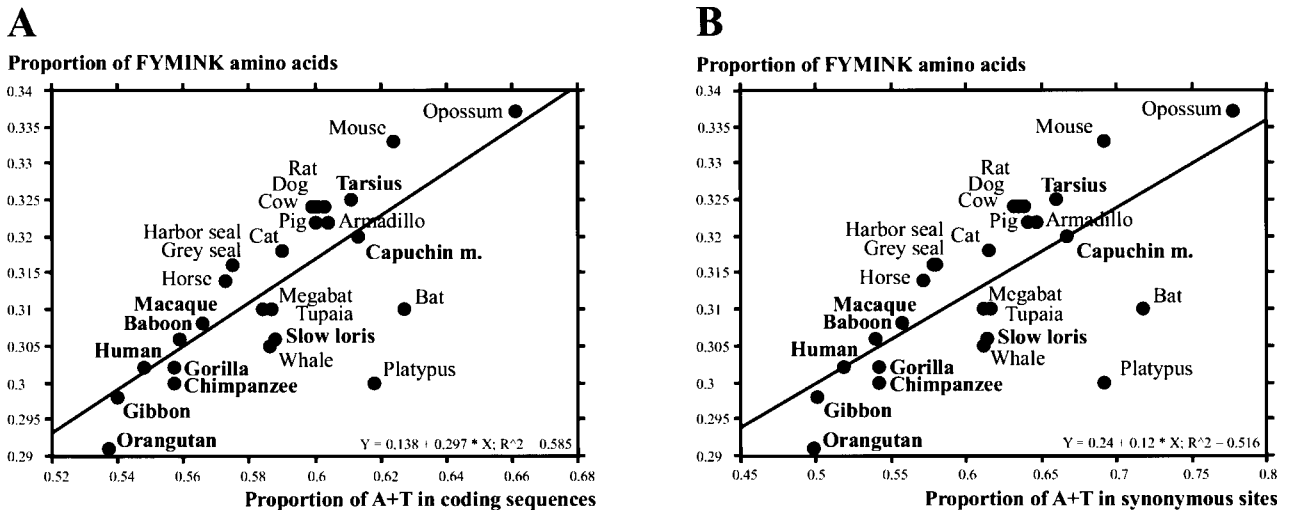

B

Proportion of FYMINK amino acids

FIG. 4.—Correlation between mean nucleotide content and AA content of all 12 H-strand–encoded mt proteins for 26 mammals. The total content of A+T-rich AAs (FYMINK) increases significantly with increasing A+T content. Note the separated position of Tarsius compared with other primates. (*A*) Proportion of A+T content, including the first, second, and third triplet position of coding sequences. (*B*) Proportion of A+T content of synonymous sites of protein-coding sequences.

Comparison with mt Pseudogenes

To determine whether the observed changes in the nucleotide composition along the lineage leading to higher primates is a general phenomenon for genomic DNA or restricted to mtDNA, we analyzed the nucleotide composition of nuclear-located mt pseudogenes that originated by a transfer of mt cytochrome *b* information into nuclear DNA during primate evolution (fig. 5). We chose mt pseudogenes described in humans, for which
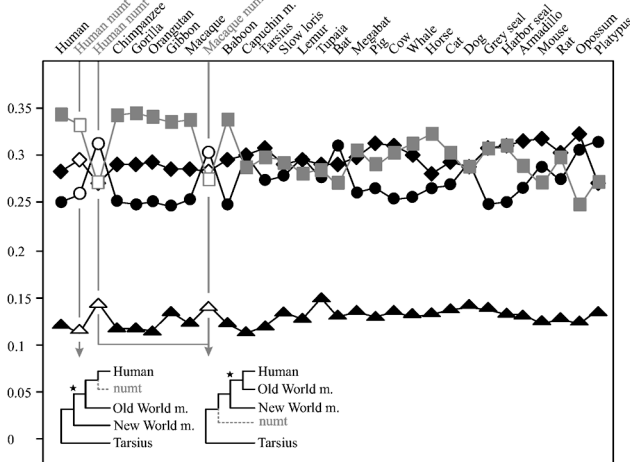

FIG. 5.—Nucleotide frequencies of 27 complete mammalian mt cytochrome *b* sequences and three corresponding nuclear pseudogenes (nuclear integrations of mtDNA [numt], open symbols). The adenosine base frequency is represented by rhombs, cytosine by gray rectangles, guanine by triangles, and thymine by circles. The cladograms display the time interval for the two integration scenarios. Left: transposition of a cytochrome *b* copy into the nucleus during the divergence of the Hominoidea and after the potential shift of nucleotide composition (labeled by an asterisk). Right: transposition of a cytochrome *b* sequence before the New World monkeys split off. Note that essentially the same results were obtained from analyses of 10 further partial cytochrome *b*-like sequences representing both transposition intervals (data not shown).

the time of transfer to the nucleus was estimated by phylogenetic tree reconstructions (data not shown). On the basis of the nuclear-flanking sequences derived from the information available from humans, we amplified and sequenced one pseudogene in rhesus macaques. Thus, the presence or absence status corroborates a time of transposition to the nucleus before the divergence of the Hominoidea. Compared with genuine, hominoid mtDNA, no obvious difference in the composition of the pseudogenes could be detected that transposed during the divergence of the Hominoidea and thus after the point in time of rate-acceleration. Because of the relatively fossilized character of nuclear integrations of mtDNA (Perna and Kocher 1996; Sunnucks and Hales 1996), we conclude that the overall base composition of mtDNA, as it existed at the time of transfer, is conserved in the nucleus. In contrast, the nuclear pseudogene detectable in the rhesus monkey that translocated before the divergence of the Anthropoidea exhibits a clearly different nucleotide composition. It can therefore be assumed that this transposition took place before the compositional shift reached its current state that can be traced in the extant taxa, and that the ancestral mt nucleotide composition was preserved. Also it can be concluded that the compositional shift is restricted to the mtDNA.

Discussion

Unequal rate effects confound reconstructions of phylogenetic trees from DNA and protein sequences (Lake 1994). The implication of this has been termed long-branch attraction (for details see Page and Holmes 1998, pp. 191–193). Here we refer to the short-branch relatives of long-branch taxa and the molecular forces that cause what we term long-branch rejection.

The intraordinal relationships of living primates, particularly the phylogenetic position of tarsiers, has

been a controversial issue for many years. Nuclear molecular data (Goodman et al. 1998, Page and Goodman 2001) and orthologous integrations of short interspersed nuclear elements (SINEs) in tarsiers and anthropoids (Schmitz, Ohme, and Zischler 2001) have since confirmed their sister taxon relationship. However, results from mtDNA analyses based on cytochrome *b* sequences either deviated from this view by excluding tarsiers from primates (Andrews, Jermiin, and Easteal 1998) or indicated only a marginal affiliation of tarsiers with anthropoids, as shown by tree reconstructions based on COII sequences (Adkins and Honeycutt 1994). By performing a more comprehensive phylogenetic analysis, including all H-strand–encoded AAs, only the MP analysis results in a sister taxon relationship between Tarsius and anthropoids (fig. 1*A*), whereas ML- and distance-based phylogenetic reconstructions consistently failed to give support for this topology. This indicates that whatever drives tarsiers out from their canonical place, as observed in most phylogenetic reconstructions, is spread throughout the mt genome with varying intensity.

Obviously, a long branch leading to higher primates can be detected in all reconstructions. An acceleration of the substitution rate in anthropoids is independently reported for COII, cytochrome *b,* and COI (Adkins, Honeycutt, and Disotell 1996; Andrews, Jermiin, and Easteal 1998; Andrews and Easteal 2000). We propose that the long branch leading to anthropoids, combined with a successive splitting of strepsirrhines and tarsiers taking place in a short time interval, and the relatively long terminal branches are responsible for the well-supported association of slow loris and Tarsius in a monophyletic group (fig. 1*A* and *B*). Moreover, this can even result in a separation of Tarsius and slow loris from primates (fig. 1*C*).

A key question now is what constitutes the driving force for the acceleration of the molecular clock in anthropoids. Adkins, Honeycutt, and Disotell (1996) and also Andrews, Jermiin, and Easteal (1998) suggested that an episode of adaptive evolution might have taken place on the lineage to higher primates. They concluded this from observing a higher relative rate of nonsynonymous to synonymous substitutions. Gillespie (1984) assumed that in mammals, in general, the molecular clock at nonsynonymous sites is episodic, with periods of bursts of substitutions as genes adapt to environmental changes.

We performed relative rate tests for the concatenated 12 H-strand–coding genes. A highly significant acceleration of nonsynonymous substitutions on the branch leading to higher primates was detected, in line with Adkins, Honeycutt, and Disotell (1996) and Andrews, Jermiin, and Easteal (1998). In contrast, it appears that the rate of synonymous substitutions is not affected. However, synonymous substitutions are prone to saturation effects overriding the real course of substitutions that took place during evolution. This is particularly pronounced in mtDNA because of its generally high rate of nucleotide substitutions (Brown et al. 1982). We therefore restricted the analysis of synonymous changes to transversions per fourfold degenerate sites,

which are thought to accumulate at a slower rate. Again, the relative rate test failed to find significant synonymous changes in the lineage leading to higher primates. These results would give strong evidence for an episode of Darwinian selection at the molecular level having an effect on all mt genes. Because none of the species analyzed showed any significant ad hoc changes in the synonymous rate, we performed a more elaborated test. For this, we arranged all possible sets of two taxa into a trio, together with their closest outgroup (data not shown). Results showed that only human and common chimpanzee when paired with gorilla, the trio constellation comprising the closest relatives in our survey, differ significantly in their synonymous substitution rate (table 2). We were thus confident that the silent sites are highly saturated, rendering the relative rate test not applicable.

Given that positive selection is not the driving force behind the acceleration of substitutions in higher primates, we propose a more generally increased mutation rate that is restricted to anthropoids.

To figure out if there is a directional pattern of mutation pressure, we compared the nucleotide composition of the terminal taxa investigated and found an obvious change on the branch leading to higher primates. Also the ancestral sequences of the MRCAs were estimated and compared to get an impression about the course of the compositional changes during primate evolution (fig. 2). The substitutions are characterized by a C accumulation in anthropoids accompanied by a decrease of A and T. A gradual shift can be deduced by taking the information of the computed compositions, as obtained from the MRCAs, into account. From this, it is apparent that catarrhines underwent a shift in a direction opposite to that of the general tendency of capuchin monkey, Tarsius, and slow loris. The overall pattern gives a first, more detailed impression about the compositional plasticity of the mt genomes in primates. All regions, coding and noncoding, constant and variable rRNA domains are affected, whereas the impact is most pronounced at selectively neutral sites (fig. 2). In line with our findings, Singer and Hickey (2000) claimed that variation in the base composition is usually mainly observed at synonymous codon positions.

If, however, both synonymous and nonsynonymous changes are affected by the variation of the base composition, Singer and Hickey (2000) expected proteins to change their AA composition, depending on the underlying nucleotide bias. Multivariate analysis of the AA usage of nonprimates and primates shows that higher primates have a higher affinity to G+C-based AAs, whereas Tarsius and slow loris tend to use nonprimate-like, distinctively more A+T-based AAs (fig. 3). In order to demonstrate that the AA bias in higher primates is the result of a nucleotide bias and not vice versa, we performed correlation analyses of AA usage and nucleotide composition for all codon positions and as a contrast, for third (silent) codon positions only (fig. 4, www.molbiolevol.org). We found an overall highly significant correlation between nucleotide bias and AA bias. Furthermore, the correlation between codon usage

and the third codon position was also highly significant. Thus, these results give further evidence that adaptive evolution is not the driving force for the changes observed because positive selection should not affect silent nucleotide positions, and the differences in AA composition are driven by the nucleotide bias. The magnitude of directional mutation pressure, as the cause of the compositional plasticity, will be fixed at a dynamic equilibrium among directional mutation pressure, the G+C content, and selective constraints (Sueoka 1992). However, this does not preclude any other kind of selection, including a positive selection for certain functional advantages, becoming effective consecutively. According to this, functional changes in the mt cytochrome *c* and cytochrome *c* oxidase interactions, as identified by kinetic studies (Osheroff et al. 1983), could initially be caused by directional mutation pressure and then be formed by positive selection (e.g., compensatory). Darwinian selection at the molecular level is characterized by selective changes of few functionally significant characters and is often expressed in independent lineages, as shown for the parallel evolution of lysozymes in colobine monkeys and ruminants (Sharp 1997). Such changes are hard to detect by exclusively analyzing the ratio of nonsynonymous versus synonymous substitutions. For cytochrome *b* sequences, Andrews, Jermiin, and Easteal (1998) identified six of 26 AA changes along the branch leading to anthropoid primates as potential candidates for functional alterations. However, the possible functional changes are not expressed in any other mammalian lineage analyzed to date. Furthermore, functional interactions between mt and nuclear genes could be affected by directional mutation pressure in mitochondria. Wu et al. (1997) reported evidence for positive selection of the cytochrome *c* oxidase subunit IV in anthropoid primates. We propose that this nuclear-encoded subunit of the mt multienzyme complex is forced to compensatorily follow the evolutionary peculiarities of the mtDNA. One might argue that the directional mutation pressure observed is not restricted to mtDNA; and therefore, it is independently expressed in mt and nuclear DNA. To address this issue, we compared mt pseudogenes detectable in higher primates: on the one hand, we analyzed the composition of pseudogenes that are assumed to have been transferred to the nucleus before the nucleotide composition changed and on the other hand, pseudogenes that transferred after this compositional shift took place. In higher primates, as an example, we found a similar nucleotide composition, as compared with mtDNA of Tarsius and nonprimates, in pseudogenes for cytochrome *b* that transferred to the nucleus before the nucleotide composition changed. In contrast, the nucleotide composition of pseudogenes, which transposed after the nucleotide content shifted, was similar to the overall composition of higher primates (fig. 5). This strongly indicates that the observed rate acceleration is a characteristic feature of the analyzed anthropoid mtDNAs. Comparison of additional genes and partial pseudogenes (data not shown) which probably reside in different genomic regions that could exhibit different A+T-content, revealed the same com-

positional characteristics in various nuclear integrations of mtDNA that transposed to the nucleus in the same time range. It is, therefore, unlikely, that postintegration substitutions drastically change the overall composition of a mtDNA-like sequence by a possible adoption of the mutational mode of the adjacent nuclear regions. Taking the pseudogene data together with the estimated sequences of the MRCAs leads to the emergence of a consistent picture about the compositional history of mtDNA in primate evolution.

Conclusions

We provide evidence, based on complete mtDNA sequences of a taxonomically broad sample, that directional mutation pressure rather than positive selection is responsible for the accelerated substitution rate in anthropoids, thus combined with a short time interval of splitting of strepsirrhines and tarsiers causing phylogenetic misinterpretations based on mtDNA. Almost all phylogenetic reconstructions therefore wrongly indicate the absence of a close relationship of Tarsius to the Anthropoidea. The overall tendency of the accelerated substitution rate for substituting A and T with C speeds up the process of saturation, particularly for synonymous transversions, leading in turn to misinterpretations of the relative rate tests. Comparisons of genuine mtDNA sequences to mt pseudogenes traced the target of the modifying forces back to the mt genome. Because of the relatively fossilized character of such nuclear mtDNA-like pseudogenes, these integrations might have conserved the mt base composition as it had existed when the transposition occurred.

On the basis of the full-length mt sequences that cover all primate infraorders, it is obvious that rate accelerations have occurred several times and independently on different evolutionary lineages in this eutherian order, as proposed for Old World and New World monkeys (Adkins, Honeycutt, and Disotell 1996) and discussed herein. Partial mtDNA sequences hint toward an even higher compositional plasticity of primate mt genomes. However, more sequence comparisons of complete mtDNA will yield more exact data on the extent and independent occurrence of compositional shifts in different primate lineages.

Given the broad taxonomic sample, as it exists to date in mtDNA data sets for the order of primates, the full picture of compositional plasticity of mtDNA in other eutherian orders is probably not yet obtained. It has to be noted that, as exemplified for the infraordinal phylogenetic relationships of primates, this compositional plasticity, boosted by an inadequate taxonomic sampling, might lead to erroneous tree reconstructions. Moreover, attempts to date branching events on the basis of mtDNA comparisons might become confounded too.

The causes of the compositional shift remain speculative. Our current understanding of possible mutational mechanisms such as DNA damage and repair and mutator-type changes in the replication process, including unbalanced nucleotide pools during replication or changes in the misincorporation behavior of the mt γ-

polymerase are still too limited to allow meaningful interpretations about how and to what extent single factors influence the nucleotide compositional plasticity of primate mtDNA.

Acknowledgments

LITERATURE CITED

ADKINS, R. M., and R. L. HONEYCUTT. 1994. Evolution of the primate cytochrome *c* oxidase subunit II gene. J. Mol. Evol. **38**:215–231.

ADKINS, R. M., R. L. HONEYCUTT, and T. R. DISOTELL. 1996. Evolution of eutherian cytochrome *c* oxidase subunit II: heterogeneous rates of protein evolution and altered interaction with cytochrome *c*. Mol. Biol. Evol. **13**:1393–1404.

ANDERSON, S., M. H. DE BRUIJN, A. R. COULSON, I. C. EPERON, F. SANGER, and I. G. YOUNG. 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. **156**:683–717.

ANDREWS, T. D., and S. EASTEAL. 2000. Evolutionary rate acceleration of cytochrome *c* oxidase subunit I in simian primates. J. Mol. Evol. **50**:562–568.

ANDREWS, T. D., L. S. JERMIIN, and S. EASTEAL. 1998. Accelerated evolution of cytochrome *b* in simian primates: adaptive evolution in concert with other mitochondrial proteins? J. Mol. Evol. **47**:249–257.

ARNASON, U., A. GULLBERG, and A. JANKE. 1997. Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenarthra (Edentata) and Ferungulates. Mol. Biol. Evol. **14**:762–768.

———. 1998. Molecular timing of primate divergences as estimated by two nonprimate calibration points. J. Mol. Evol. **47**:718–727.

ARNASON, U., A. GULLBERG, E. JOHNSSON, and C. LEDJE. 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., A. GULLBERG, and B. WIDEGREN. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. J. Mol. Evol. **33**:556–568.

ARNASON, U., A. GULLBERG, and X. XU. 1996. A complete mitochondrial DNA molecule of the white-handed gibbon, *Hylobates lar*, and comparison among individual mitochondrial genes of all hominoid genera. Hereditas **124**:185–189.

ARNASON, U., and E. JOHNSSON. 1992. The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. J. Mol. Evol. **34**:493–505.

ARNASON, U., X. XU, and A. GULLBERG. 1996. Comparison between the complete mitochondrial DNA sequences of *Homo* and the common chimpanzee based on nonchimeric sequences. J. Mol. Evol. **42**:145–152.

BIBB, M. J., R. A. VAN ETTEN, C. T. WRIGHT, M. W. WALBERG, and D. A. CLAYTON. 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell **26**:167–180.

BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17**:540–552.

FELSENSTEIN, J. 1995. PHYLIP: phylogeny inference package. Version 3.5. Distributed by the author. Department of Genetics, University of Washington, Seattle.

FLEAGLE, J. G. 1999. Primate adaptation and evolution. Academic Press, San Diego.

FOSTER, P. G., L. S. JERMIIN, and D. A. HICKEY. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J. Mol. Evol. **44**:282–288.

GADALETA, G., G. PEPE, G. DE CANDIA, C. QUAGLIARIELLO, E. SBISA, and C. SACCONE. 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.

GILLESPIE, J. H. 1984. The molecular clock may be an episodic clock. Proc. Natl. Acad. Sci. USA **81**:8009–8013.

GOODMAN, M., C. A. PORTER, J. CZELUSNIAK, S. L. PAGE, H. SCHNEIDER, J. SHOSHANI, G. GUNNELL, and C. P. GROVES. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **9**:585–598.

GUTELL, R. R., and G. E. FOX. 1988. A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res. **16**:r175–r313.

HIXON, J. E., and W. M. BROWN. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: phylogenetic implications. Mol. Biol. Evol. **3**:1–18.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92**:532–536.

JANKE, A., G. FELDMAIER-FUCHS, W. K. THOMAS, A. VON HAESELER, and S. PÄÄBO. 1994. The marsupial mitochondrial genome and the evolution of placental mammals. Genetics **137**:243–256.

JANKE, A., N. J. GEMMELL, G. FELDMAIER-FUCHS, A. VON HAESELER, and S. PÄÄBO. 1996. The mitochondrial genome of a monotreme—the platypus (Ornithorhynchus anatinus). J. Mol. Evol. **42**:153–159.

KIM, K. S., S. E. LEE, H. W. JEONG, and J. H. HA. 1998. The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. Mol. Phylogenet. Evol. **10**:210–220.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29**:170–179.

LAKE, J. A. 1994. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. Proc. Natl. Acad. Sci. USA **91**:1455–1459.

LOPEZ, J. V., S. CEVARIO, and S. O'BRIEN. 1996. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (numt) in the nuclear genome. Genomics **33**:229–246.

NIKAIDO, M., M. HARADA, Y. CAO, M. HASEGAWA, and N. OKADA. 2000. Monophyletic origin of the order chiroptera and its phylogenetic position among mammalia, as inferred from the complete sequence of the mitochondrial DNA of a Japanese megabat, the Ryukyu flying fox (*Pteropus dasymallus*). J. Mol. Evol. **51**:318–328.

OSHEROFF, N., S. H. SPECK, E. MARGOLIASH, E. C. VEERMAN, J. WILMS, B. W. KÖNIG, and A. O. MUIJSERS. 1983. The

reaction of primate cytochromes *c* with cytochrome *c* oxidase. J. Biol. Chem. **258**:5731–5738.

PAGE, R. D., and E. C. HOLMES. 1998. Molecular evolution: a phylogenetic approach. Blackwell Science, Oxford.

PAGE, S. L., and M. GOODMAN. 2001. Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Mol. Phylogenet. Evol. **18**:14–25.

PERNA, N. T., and T. D. KOCHER. 1996. Mitochondrial DNA: molecular fossils in the nucleus. Curr. Biol. **6**:128–129.

PUMO, D. E., P. S. FINAMORE, W. R. FRANEK, C. J. PHILLIPS, S. TARZAMI, and D. BALZARANO. 1998. Complete mitochondrial genome of a neotropical fruit bat, *Artibeus jamaicensis*, and a new hypothesis of the relationships of bats to other eutherian mammals. J. Mol. Evol. **47**:709–717.

ROBINSON, M., M. GOUY, C. GAUTIER, and D. MOUCHIROUD. 1998. Sensitivity of the relative-rate test to taxonomic sampling. Mol. Biol. Evol. **15**:1091–1098.

SBISA, E., F. TANZARIELLO, A. REYES, G. PESOLE, and C. SACCONE. 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene **205**:125–140.

SCHMITZ, J., M. OHME, and H. ZISCHLER. 2000. The complete mitochondrial genome of *Tupaia belangeri* and the phylogenetic affiliation of Scandentia to other eutherian orders. Mol. Biol. Evol. **17**:1334–1343.

———. 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics **157**:777–784.

SHARP, P. M. 1997. In search of molecular darwinism. Nature **385**:111–112.

SINGER, G. A. C., and D. A. HICKEY. 2000. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. Mol. Biol. Evol. **17**:1581–1588.

SPRINZL, M., C. HORN, M. BROWN, A. IOUDOVITCH, and S. STEINBERG. 1998. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. **26**:148–153.

STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

SUEOKA, N. 1992. Directional mutation pressure, selective constraints, and genetic equilibria. J. Mol. Evol. **34**:95–114.

SUNNUCKS, P., and D. F. HALES. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Mol. Biol. Evol. **13**:510–524.

SWOFFORD, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**:4876–4882.

URSING, B. M., and U. ARNASON. 1998. The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). J. Mol. Evol. **47**:302–306.

WADDELL, P. J., Y. CAO, J. HAUF, and M. HASEGAWA. 1999. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. Syst. Biol. **48**:31–53.

WU, W., M. GOODMAN, M. I. LOMAX, and L. I. GROSSMAN. 1997. Molecular evolution of cytochrome *c* oxidase subunit IV: evidence for positive selection in simian primates. J. Mol. Evol. **44**:477–491.

XU, X., and U. ARNASON. 1994. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. Gene **148**:357–362.

ZHANG, D.-X., and G. M. HEWITT. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol. Evol. **11**:247–251.

ZUKER, M., D. H. MATHEWS, and D. H. TURNER. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. Pp. 11–43 *in* J. BARCISZEWSKI and B. F. CLARK, eds. RNA biochemistry and biotechnology. Kluwer Academic Publishers, Dordrecht.