The Complete Mitochondrial Genome of *Tupaia belangeri* and the Phylogenetic Affiliation of Scandentia to Other Eutherian Orders

Jürgen Schmitz, Martina Ohme, and Hans Zischler

Primate Genetics, German Primate Center, Göttingen, Germany

The complete mitochondrial genome of *Tupaia belangeri,* a representative of the eutherian order Scandentia, was determined and compared with full-length mitochondrial sequences of other eutherian orders described to date. The complete mitochondrial genome is 16,754 nt in length, with no obvious deviation from the general organization of the mammalian mitochondrial genome. Thus, features such as start codon usage, incomplete stop codons, and overlapping coding regions, as well as the presence of tandem repeats in the control region, are within the range of mammalian mitochondrial (mt) DNA variation. To address the question of a possible close phylogenetic relationship between primates and Tupaia, the evolutionary affinities among primates, Tupaia and bats as representatives of the Archonta superorder, ferungulates, guinea pigs, armadillos, rats, mice, and hedgehogs were examined on the basis of the complete mitochondrial DNA sequences. The opossum sequence was used as an outgroup. The trees, estimated from 12 concatenated genes encoded on the mitochondrial H-strand, add further molecular evidence against an Archonta monophyly. With the new data described in this paper, most of both the mitochondrial and the nuclear data point away from Scandentia as the closest extant relatives to primates. Instead, the complete mitochondrial data support a clustering of Scandentia with Lagomorpha connecting to the branch leading to ferungulates. This closer phylogenetic relationship of Tupaia to rabbits than to primates first received support from several analyses of nuclear and partial mitochondrial DNA data sets. Given that short sequences are of limited use in determining deep mammalian relationships, the partial mitochondrial data available to date support this hypothesis only tentatively. Our complete mitochondrial genome data therefore add considerably more evidence in support of this hypothesis.

Introduction

There is considerable debate about which eutherian order shares a common ancestry with primates exclusive of all other mammalians. The concept of the superorder Archonta, in which, traditionally, primates and the hypothetical sister taxa bats (Chiroptera), flying lemurs (Dermoptera), and tree shrews (Scandentia) are grouped together in various constellations, is based mainly on anatomical evidence obtained from extant species (Novacek 1992). Neither paleontological data (Kay, Thorington, and Houde 1990) nor most interpretations of molecular data (e.g., Bailey, Slightom, and Goodman 1992; Graur, Duret, and Gouy 1996; Porter, Goodman, and Stanhope 1996) support this concept.

Among other molecular evidence, tree reconstructions obtained from the complete mitochondrial DNA sequences (Pumo et al. 1998) reveal phylogenetic affiliations of bats to ferungulates (carnivores, perissodactyls, artiodactyls, and cetaceans). Moreover, the supposed close phylogenetic relationship between Dermoptera and Chiroptera, suggested by various morphological analyses, is not supported by large-scale molecular studies combining nuclear and mitochondrial sequences (Teeling et al. 2000). Therefore, it seems conceivable that numerous morphological characters formerly recognized as shared derived characters and linking bats

and flying lemurs are largely due to convergences that appeared with the evolution of gliding or flight. These findings shed doubts on the validity of the Archonta superorder concept and have reopened the question of primate affinities with the other Archonta members Dermoptera and Scandentia. In order to put the latter into an interordinal phylogenetic framework with other mammals, we sequenced the complete mitochondrial genome of *Tupaia belangeri.*

The Tupaiidae represent a small radiation currently consisting of 18 recognized extant species geographically distributed in south and southeast Asia (Martin 1990). Their uncertain phylogenetic affiliation is reflected in the various systematic classifications attributed to this taxon in the past. Before being included in the mammalian order of Scandentia, the Tupaiidae were recognized as the deepest primate split, after they were classified as being a member of the former insectivore subgroup Menotyphla (Martin 1990). This uncertainty is mainly due to the fact that Tupaia shows a complex mixture of plesiomorphic and apomorphic morphological characters (Starck 1978; Martin 1990), a lack of fossil information pertaining to the Scandentia, and the long independent evolutionary history of Scandentia (Starck 1978). Nevertheless, Tupaia is currently generally recognized as an intermediate between primates and other eutherian orders. Since it serves as a widely used animal model in primate- and human-related research, its exact phylogenetic position is not only of biological interest regarding itself. The study of the evolution of, for example, anatomical, physiological, life history, and karyological characters in primates thus necessitates a firm establishment of an appropriate outgroup for the mammalian order to which our species belongs.

Materials and Methods

Isolation and Amplification of Mitochondrial DNA

The mtDNA of *T. belangeri* was isolated from fresh liver samples. In order to prevent inadvertent amplifications of mitochondrial pseudogenes residing in the nuclear genome, mtDNA was enriched in the DNA preparation using the differential centrifugation procedure described in Arnason, Gullberg, and Widegren (1991). Second, longer fragments, exceeding the average length of nuclear integrations of mitochondrial DNA described to date (Zhang and Hewitt 1996), were amplified first and further processed in nested PCRs. Third, all coding sequences were carefully checked for indels and premature stops, which interfere with the synthesis of a correct gene product. The enriched mtDNA fraction was then used to amplify five overlapping 3–4-kb-long PCR fragments with conserved primers constructed on the basis of human-mouse comparisons. Versatile primers were then designed to amplify 19 overlapping fragments of about 1 kb in length in a second round of PCR starting from the long PCR fragments as templates. The same primers were also used to amplify the smaller fragments directly from the enriched mtDNA. The resulting sequences were compared with those generated as outlined above. Each final sequence constituted a consensus of a minimum of three individual clones from at least two separate PCR products.

Both the long PCR fragments and the 19 1-kb fragments were ligated into pGEM-T vector (PROMEGA) and electroporated into TOP10 cells (INVITROGEN). Plasmid sequencing was performed with universal primers using an automated LI-COR DNA sequencer 4200.

The mitochondrial sequence of *T. belangeri* is available in GenBank under the accession number AF217811.

Analyses of Sequence Data

The following complete eutherian mtDNA sequences were retrieved from the database and put in a phylogenetic context with Tupaia: human, X93334 (Arnason, Xu, and Gullberg 1996); common and pygmy chimpanzee, D38113, D38116; gorilla, D38114; orangutan, D38115 (Horai et al. 1995); gibbon, X99256 (Arnason, Gullberg, and Xu 1996); hamadryas baboon, Y18001 (Arnason, Gullberg, and Janke 1998); guinea pig, NC_000884 (D'Erchia et al. 1996); rabbit, AJ001588 (Gissi, Gullberg, and Arnason 1998); armadillo, Y11832 (Arnason, Gullberg, and Janke 1997); bat, AF061340 (Pumo et al. 1998); cow, J01394 (Anderson et al. 1982); fin whale, X61145 (Arnason, Gullberg, and Widegren 1991); pig, AJ002189 (Ursing and Arnason 1998); harbor seal, X63726 (Arnason and Johnsson 1992); gray seal, X72004 (Arnason et al. 1993); horse, X79547 (Xu and Arnason 1994); dog, U96639 (Kim et al. 1998); cat, U20753 (Lopez et al. 1996); rat, X14848 (Gadaleta et al. 1989); mouse, J01420 (Bibb et al. 1981); hedgehog, X88898 (Krettek, Gullberg, and Arnason 1995); opossum, Z29573 (Janke et al. 1994).

The phylogenetic analyses were primarily carried out on the basis of amino acid sequences. For phylogenetic analyses, we concatenated the 12 protein-coding genes encoded on the mitochondrial H-strand. The L-strand–encoded ND6 was excluded from the analyses, since its composition deviates significantly from that of the other mitochondrial genes. In addition, the rRNA-specifying regions were considered for the phylogenetic analyses.

The alignments of amino acid and nucleotide sequences were carried out by applying CLUSTAL X (Thompson et al. 1997). The alignment proposed by D'Erchia et al. (1996) (http://www.ba.cnr.it/guinea-pig.html) was used as a profile for the newly added amino acid and nucleotide sequences. For phylogenetic analyses based on nucleotide sequences of the 12 concatenated genes, all third codon positions were excluded.

A priori tests of the data for the presence of a phylogenetic signal were carried out using the likelihood-mapping option implemented in PUZZLE, version 4.0.2 (Strimmer and von Haeseler 1996, 1997). Phylogenetic reconstructions were carried out using three methods: maximum parsimony (MP), included in PAUP* version 4.0b4a (Swofford 1999), neighbor-joining (NJ) analyses (Saitou and Nei 1987) using PHYLIP, version 3.573c (Felsenstein 1995), and maximum likelihood (ML) as implemented in PUZZLE, version 4.0.2 (Strimmer and von Haeseler 1996), for both amino acids (aa) and DNA sequences. Heuristic parsimony analyses were conducted with random taxon addition and tree bisection-reconnection (TBR) branch swapping. In searches, all most-parsimonious trees were stored. Distance corrections for the NJ analysis of aa sequences were done on the basis of the Dayhoff PAM matrix. The ML analyses on the protein level were carried out assuming the mtREV24 (Adachi and Hasegawa 1996) model of sequence evolution either assuming no rate variation or approximating a gamma distribution of rates across sites by introducing four and eight rate categories. The respective gamma distribution parameter alpha was estimated from the data set, as was the frequency of the amino acids.

For ML reconstructions of the protein-coding nucleotide sequences (first and second codon positions), the HKY model of sequence evolution with rate heterogeneity across sites was assumed (Hasegawa, Kishino, and Yano 1985). The respective NJ analyses were carried out both with ML distance corrections as implemented in PHYLIP and on the basis of Markov distance corrections (Saccone et al. 1990).

Support of internal branches either was determined by bootstrap analysis (MP and NJ) performed with 1,000 replications or was indicated by the ML quartet puzzling support values (1,000 puzzling steps). Majority rule consensus trees were calculated using the CONSENSE program of the PHYLIP package.

Structural Analyses of RNAs

The secondary structures of the tRNAs, as well as those of the ribosomal RNAs, were reconstructed on the basis of the compilation of Sprinzl et al. (1998) (http://www.uni-bayreuth.de/departments/biochemie/trna), and

Table 1
Organization of the *Tupaia belangeri* Mitochondial Genome

Gene[a]	From	To	Size	Start Codon	Stop Codon[b]	Spacer/Overlap
tRNA-Phe	1	66	66			
12S rRNA	67	1,014	948			
tRNA-Val	1,015	1,080	66			
16S rRNA	1,081	2,652	1,572			
tRNA-Leu (UUR)	2,653	2,727	75			2-base spacer
NADH 1	2,730	3,686	957	ATG	Tag	2-base overlap
tRNA-Ile	3,685	3,753	69			3-base overlap
tRNA-Gln (L)	3,751	3,821	71			1-base overlap
tRNA-Met	3,821	3,889	69			
NADH 2	3,890	4,933	1,044	ATC	Tag	2-base overlap
tRNA-Trp	4,932	4,999	68			5-base spacer
tRNA-Ala (L)	5,005	5,074	70			1-base spacer
tRNA-Asn (L)	5,076	5,148	73			
Origin-L (L)	5,149	5,181	33			
tRNA-Cys (L)	5,182	5,249	68			
tRNA-Tyr (L)	5,250	5,316	67			
COI	5,317	6,858	1,542	ATG	TAA	2-base spacer
tRNA-Ser (UCN) (L)	6,861	6,929	69			4-base spacer
tRNA-Asp	6,934	7,002	69			
COII	7,003	7,686	684	ATG	TAG	
tRNA-Lys	7,687	7,752	66			2-base spacer
ATPase8	7,755	7,958	204	ATG	TAA	43-base overlap
ATPase6	7,916	8,596	681	ATG	TAa	1-base overlap
COIII	8,596	9,381	786	ATG	TAa	1-base overlap
tRNA-Gly	9,381	9,446	66			
NADH 3	9,447	9,793	347	ATA	TA–	
tRNA-Arg	9,794	9,859	66			1-base spacer
NADH 4L	9,861	10,157	297	ATG	TAA	7-base overlap
NADH 4	10,151	11,527	1,377	ATG	TA–	1-base spacer
tRNA-His	11,529	11,597	69			
tRNA-Ser (AGY)	11,598	11,655	58			1-base overlap
tRNA-Leu (CUN)	11,655	11,724	70			
NADH 5	11,725	13,536	1,812	ATC	TAA	1-base spacer
NADH 6 (L)	13,538	14,059	522	ATG	TAG	
tRNA-Glu (L)	14,060	14,127	68			3-base spacer
Cytb	14,131	15,270	1,140	ATG	TAG	
tRNA-Thr	15,271	15,335	65			2-base spacer
tRNA-Pro (L)	15,338	15,404	67			
Control region	15,405	16,754	1,350			

[a] "(L)" indicates a gene encoded on the L-strand.
[b] Lowercase letters indicate possible overlap with the beginning of the following gene.

the secondary structure models of 12S and 16S rRNA were from Springer and Douzery (1996) and Gutell and Fox (1988), respectively.

Sequence Analyses of the Control Region

Conserved regions and domains in the control region were identified by comparing the Tupaia sequence with a consensus sequence obtained from a multialignment of published conserved control region domains of different mammals (http://www.ba.cnr.it/dloop.html) (Sbisa et al. 1997). The sequence of the complete control region was checked for the presence of tandem repetitive elements using the Tandem Repeats Finder software available at http://c3.biomath.mssm.edu/trf.html (Benson 1999).

Results
Genome Content and Organization

The mitochondrial genome of *T. belangeri* is 16,754 nt in length. The L-strand base composition is as follows: A, 32.7%; C, 26.4%; G, 14.4%; T, 26.5%. Table 1 summarizes the lengths, possible overlaps, and positions of all identified polypeptide-coding genes and RNA-specifying and noncoding regions in the Tupaia mitochondrial genome.

Protein-Coding Genes

Eleven out of 13 protein-coding genes of Tupaia initiate with a methionine start codon (ATG or ATA). The start codon of the remaining genes, NADH 2 and NADH 5, is ATC. According to Fearnley and Walker (1987), any ATN start codon may specify methionine in mammalian mitochondria.

Two protein-coding genes, NADH 3 and NADH 4, are terminated by an incomplete stop codon (TA–). It has been proposed that such incomplete stop codons are completed by a posttranscriptional addition of 3′ A residues (Ojala, Montoya, and Attardi 1981). Furthermore, it cannot be excluded that NADH 1, NADH 2, ATPase 6, and COIII also possess incomplete stop codons (table 1), since their stop codons share nucleotides with the
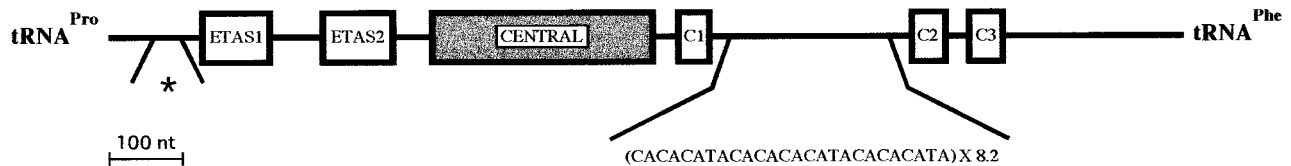
Fig. 1.—Diagrammatic representation of the Tupaia control region with flanking tRNA-specifying regions (drawn to scale). Extended termination-associated sequences (ETAS 1, ETAS 2), the central domain (CENTRAL), and the conserved sequence blocks (CSB 1, CSB 2, and CSB 3) are depicted as open boxes labeled C1, C2, and C3. The position of the repetitive consensus sequence and the number of reiterated motifs found for the individual sequenced in this study are indicated between C1 and C2. The region of 66% sequence similarity to ETAS 1 is marked with an asterisk.

adjacent start codon or tRNA-specifying regions. Compared with the other eutherian mtDNAs, Tupaia, mice, and rats lack one amino acid at the same position inside the NADH 6 gene.

RNA Genes

The sizes of the tRNAs range from 58 to 75 nt. The putative secondary structures of the 22 tRNA genes do not vary substantially from the typical mammalian patterns. Substitutions in stems were predominantly compensated by substitutions in the complementary strand. Most of the indels occurred in the dihydrouridine and TψC arms.

The 12S and 16S RNA genes are 948 and 1,572 nt in length. Corresponding to available secondary-structure models (Gutell and Fox 1988; Springer and Douzery 1996), we recognized 490 nt (12S rDNA) and 1,033 nt (16S rDNA) in single-strand conformation and 458 nt (12S rDNA) and 539 nt (16S rDNA) in base-paired regions.

Noncoding Regions

As in other mammals, the L-strand origin of replication (Ori-L) is located between the sequences specifying the Asn- and Cys-tRNA in Tupaia. With two transitions and two additional nucleotides in the loop region detectable upon comparison with the human sequence, it can be folded into the same hairpin structure outlined in Gissi, Gullberg, and Arnason (1998). No intronic spacers were found in Tupaia located between COII and the tRNA lysine and between the tRNA tyrosine and the COI gene, respectively. The latter region is variable in eutherians, with insertions up to 42 nt long in the orangutan. The absence of spacer nucleotides between COII and tRNA lysine in Tupaia is in concordance with the guinea pig sequence and in contrast to the situation in the remaining eutherian orders analyzed to date.

The mitochondrial control region of the sequenced *T. belangeri* individual is 1,350 nt in length. The typical tripartite organization in two variable domains adjacent to the Pro-tRNA (ETAS domain) and the Phe-tRNA (CSB-domain) with a central, more conserved region could be verified by aligning the Tupaia sequence to the consensus sequences described in Sbisa et al. (1997). The relative positions of the conserved blocks in the two variable domains possibly associated with the arrest and initiation of H-strand synthesis (ETAS 1, ETAS 2, CSB 1, CSB 2, and CSB 3) are diagrammatically displayed

in figure 1. The compositional analysis of the different control region domains revealed an AT preponderance in all domains of Tupaia. The two ETAS domains are separated by an AT-rich sequence of 79 nt. This means the practically contiguous ETAS domains, detectable in artiodactyls and perissodactyls, cannot be seen in Tupaia. We detected a 59-nt region 22–80 nt upstream of ETAS 1 showing 66% sequence similarity with the ETAS 1 mammalian consensus sequence published in Sbisa et al. (1997).

Adjacent to the CSB 1, a tandem repeat sequence spanning about 210 nt could be identified (its location is indicated by the sequence motif and number of reiterations in fig. 1).

Since the longest motif of the reiterated sequence is in itself composed of repetitive sequence elements, various tandem repeat motifs of either 8, 18, or 26 nt repeat length could be deduced. For this individual, the longest 26-nt motif is reiterated 8.2 times with all of the eight motifs being different from each other. The consensus sequence of the 26-nt motif was determined to be CACACATACACACACATACACACATA.

Phylogeny

For the phylogenetic analyses, we considered primarily the aa sequence of 12 concatenated protein-coding sequences encoded on the mitochondrial H-strand. The complete data set consisted of seven representatives of the primates (the human, the common chimpanzee, the pygmy chimpanzee, the gorilla, the orangutan, the gibbon, and the hamadryas baboon), one bat, and the Tupaia sequence, all of them representing members of the Archonta superorder. These taxa were phylogenetically related to the mouse, the rat, the rabbit, the guinea pig, the armadillo, the horse, the cow, the pig, the fin whale, the gray seal, the harbor seal, the dog, the cat, and the hedgehog. In general, the opossum mitochondrial sequence was used to root the phylogenetic trees.

To check for the presence of a phylogenetic signal, we conducted a likelihood-mapping analysis as implemented in PUZZLE. Therefore, we applied a gamma distribution with eight categories to consider the evolutionary rate heterogeneity across sites and assumed the mtREV24 model of sequence evolution. Including all 24 species, without grouping, a strong phylogenetic signal indicated the suitability of the data set for a phylogenetic reconstruction and ruled out a starlike evolution.

In a second step, we inferred the phylogenetic relationships of the taxa in question by a ML reconstruc-

tion (PUZZLE) based on the aa sequences and rooted with the opossum sequence. The ML analysis was performed assuming the mtREV24 model of aa evolution (Adachi and Hasegawa 1996). As a result, we obtained a tree that was not completely resolved (fig. 2A, left). In this tree, the hedgehog takes a basal position among all of the other eutherian sequences as identified by Mouchaty et al. (2000), followed by a trifurcation leading to the mouse/rat monophylum, the guinea pig, and a monophylum composed of the remaining eutherians. In the latter clade, all primates constitute a monophyletic cluster supported by the highest reliability value. The sister group to primates is represented by a clade composed of ferungulates (the cow, the pig, the fin whale, the horse, the gray seal, the harbor seal, the dog, and the cat), the bat, the armadillo, the rabbit, and Tupaia. This cluster is supported by a reliability value of 54. For Tupaia, a sister taxon relationship to the rabbit could be inferred, which is confirmed by a reliability value of 76. Both quartet puzzling support values are significantly increased upon taking only four rate categories into account or assuming no rate variation across sites (65 and 95 for the internal edge separating Tupaia from primates and 80 and 93 for the branch leading to the common ancestor of Tupaia and the rabbit).

To test the support for the internal edge of the ML tree that separates Tupaia from primates (fig. 2A, left), we performed a four-cluster likelihood-mapping analysis, implemented in PUZZLE, in which we partitioned the sequences in four disjoint groups, consisting of (1) all taxa that split off before the most recent common ancestor of Tupaia and primates (the opossum, the hedgehog, the mouse, the rat, the guinea pig; designated in the following as the outgroup), (2) the primates (the human, the common chimpanzee, the pygmy chimpanzee, the gorilla, the orangutan, the gibbon, and the hamadryas baboon), (3) Tupaia and rabbit, and (4) all taxa composing the sister clade to Tupaia/rabbit, represented by the ferungulates, the armadillo, and the bat. From the respective likelihood-mapping analyses, the strongest support (61.9% of all possible quartets) was obtained for the phylogenetic arrangements grouping primates and the composite outgroup versus Tupaia/rabbit and the remaining taxa in the data set. A second alternative grouping of primates together with the ferungulates, including the bat and the armadillo, received 22.7% support, whereas only 15.5% of all possible quartets support the topology that groups Tupaia/rabbit as a sister clade to the primates.

To corroborate the ML analyses, different phylogenetic reconstruction methods were applied to the same data set (not shown). At first, an NJ analysis on 1,000 bootstrap replicates basically recapitulated the results concerning the position of Tupaia obtained as outlined (fig. 2A, left). Thus, the cluster of Tupaia and the rabbit is supported by a bootstrap value of 54% and shows a sister taxon relationship to the armadillo and the bat/ferungulates confirmed by 99%. Moreover, the MP reconstruction (1,000 replicates) confirms the cluster of the ferungulates, the bat, the armadillo, the rabbit, and Tupaia by 79%, although the exact phylogenetic relation

among them could not be resolved. Thus a basal polytomy of Tupaia, the rabbit, the armadillo, and the bat/ferungulates clade is displayed.

With the exception of the hedgehog, the aa compositions conform to the frequencies assumed by the ML model tested in a 5% level χ^2 test at the protein level. To rule out a possibly disturbing effect on the topology of the reconstructed tree, we excluded the hedgehog from the ML, MP, and NJ analyses. In addition, the guinea pig and the rabbit, both species for which the phylogenetic positions are difficult to assign on the basis of complete mitochondrial sequences (Mouchaty et al. 2000; Gissi, Gullberg, and Arnason 1998), were removed. An alteration of the tree topologies of this pruned data set (fig. 2B) compared with the extended one (fig. 2A) could not be observed, but rather substantial increases in the support values for single nodes were generated: The clade comprising (((bat, ferungulates), armadillo), Tupaia) was supported by 58%, 78%, and 98% (ML with eight rate categories, with four rate categories, and without rate categories), 87% (MP), and 99% (NJ).

To test whether this topology received significant support from the data, the log L values of a user-defined tree with Tupaia set as the sister group of primates (log $L = -53,097.46$) and the tree shown in figure 2A (left) (log $L = -53,081.87$) were compared, revealing a likelihood difference of 15.59 ± 12.40 (SE). Thus, for the aa sequence analysis, the grouping of Tupaia and the rabbit was supported, but Tupaia as a sister group to primates could not be rejected at the 5% level of significance (Kishino and Hasegawa 1989). We therefore reduced the phylogenetic noise by pruning the data set to the taxa opossum, Tupaia, bat, cow, and three representatives of the primates, which resulted in the ML topology (opossum, ((Tupaia, (bat, cow)), (baboon, (gibbon, human)))), with a log L value of $-26,168.89$ and an MP tree length of 3,318 steps. For this sample of taxa, the MP tree length difference (24 ± 10.09 [SD]), as well as the likelihood difference from assuming a uniform substitution rate (35.17 ± 16.95 [SE]) of the pruned tree topology shown above compared with the tree placing Tupaia as a sister taxon to primates, rejects significantly a close relationship between primates and Tupaia (5% significance level of Kishino and Hasegawa 1989).

Concatenated nucleotide sequences of protein-coding genes were used to reconstruct the phylogeny shown in figure 2 (right). In an NJ analysis, we used the genetic distances calculated from the first and second codon positions with the stationary Markov model (Saccone et al. 1990). Tupaia is placed together with the rabbit and the armadillo at the branch leading to the ferungulates/the bat with 77% (fig. 2A, right) and 94% (fig. 2B, right) support. The ML reconstruction remains widely unresolved (not shown). Concerning the phylogenetic position of Tupaia, the pruned MP tree confirmed the tree shown in figure 2B (right) and indicates 90% support for the Tupaia and ferungulates/bat cluster (not shown).

In order to extract additional phylogenetic information, further examinations were carried out on the
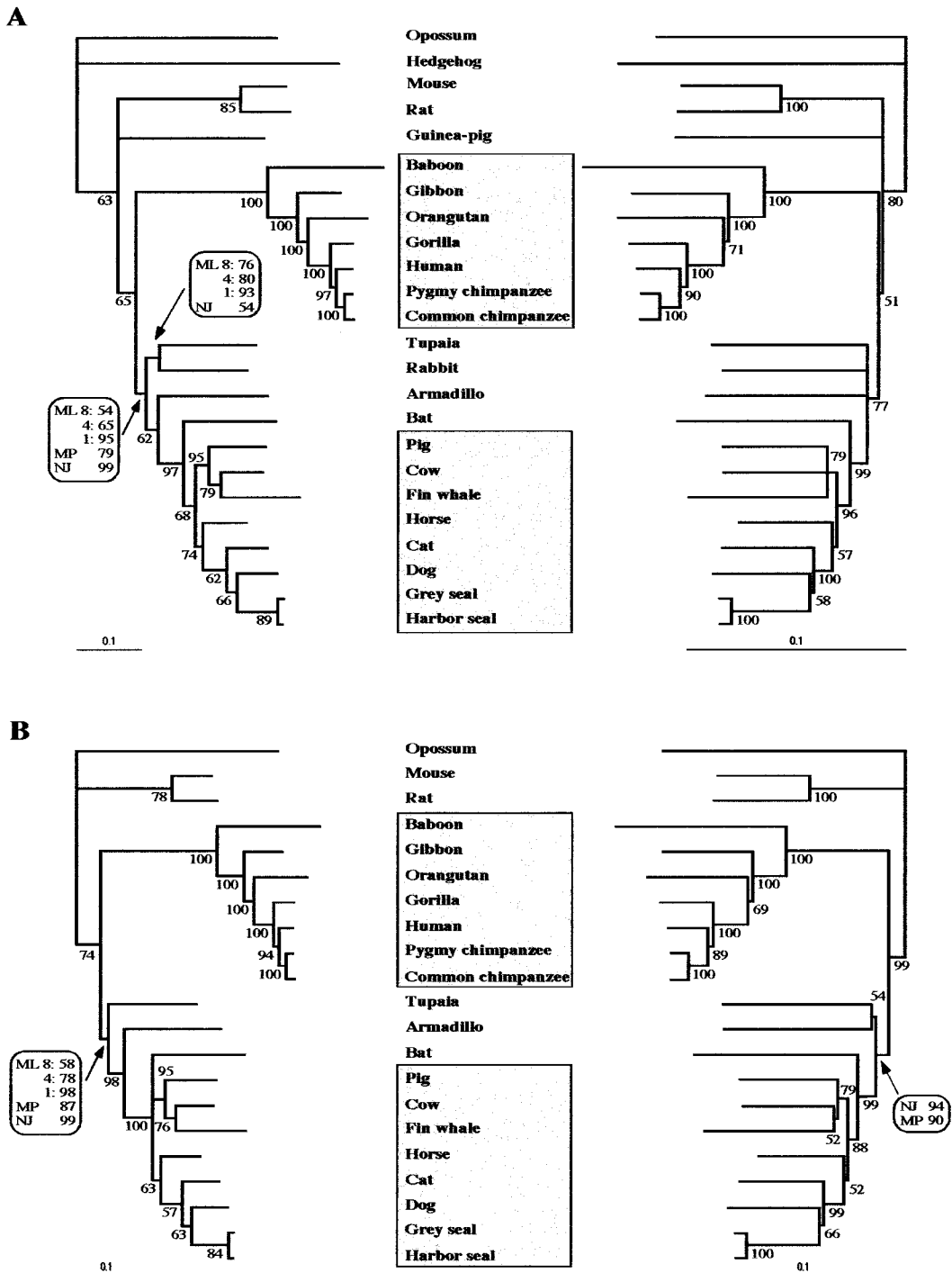
Fig. 2.—Tree reconstructions of the mammalian relationships supported by the 12 H-strand protein-coding genes, (A) including and (B) excluding the guinea pig, the rabbit, and the hedgehog (see text). The metatherian opossum was used as an outgroup. Bootstrap or quartet puzzling support values are indicated below the corresponding nodes. Nodes supported by less than 50% were collapsed into polytomies. Alternative support values revealed by maximum likelihood (ML) performed with eight rate categories, with four rate categories, and without different rate categories (ML 8, ML 4, and ML 1), neighbor-joining (NJ), and maximum-parsimony (MP) reconstructions are depicted in boxes. The bars indicate 0.1 amino acid or nucleotide substitutions per position, respectively. On the left: majority-rule consensus tree obtained by the ML method using the PUZZLE program (Strimmer and von Haeseler 1996) on the amino acid sequences. Support values were based on 1,000 puzzling steps. The ML branch lengths are proportional to substitutions of amino acids per site. On the right: majority-rule consensus tree obtained by the NJ method (Saitou and Nei 1987) of the first and second codon positions of the 12 H-strand proteins using the distance matrix as calculated with the stationary Markov model (Saccone et al. 1990). Branch lengths represent nucleotide substitutions per site.

basis of 12S and 16S rDNA sequences. An initial like-lihood-mapping analysis (without grouping) of the con-catenated 12S rRNA and 16S rRNA genes indicated a high degree of starlikeness in the complete data set, as well as in the separated genes. This points to a low phylogenetic resolution power, which was confirmed in subsequent tree reconstructions (MP, ML, and NJ). In order to overcome a probable high degree of homopla-sious exchanges in the data set, we considered the vary-ing patterns of nucleotide substitutions in different re-gions of the 12S rRNA and 16S rDNA–specifying se-quences. For this purpose, the 12S rRNA and 16S rDNA were split into two separate subsets relative to their sec-ondary structures, with regions either adopting a single-strand conformation (loops) or being potentially base-paired (stems). Compared with all other partitions, the likelihood-mapping analysis revealed the strongest phy-logenetic signal in the stem region data set of the com-bined 12S/16S rDNA. Therefore, 12S/16S rDNA stems were used in a following tree reconstruction. As a result, a sister group relationship between Tupaia and the rab-bit, confirmed by 77% (ML), 55% (MP), and 69% (NJ) support, could be obtained (data not shown).

Discussion
Organization of the Mitochondrial Genome

Regarding the gene content and arrangement in the mitochondrial genome, the Tupaia sequence conforms to the mtDNAs of other eutherians. Features like incom-plete stop codons, overlapping coding regions, and dif-ferent start codons lie within the range of mammalian mtDNA variation. All tRNA-specifying regions of Tu-paia were folded and compared with the respective tRNA structures of other eutherians. From this, we con-clude that all of them are fully functional.

The major noncoding region of the mitochondrial genome, the control region, exhibits the typical tripartite structure with a central domain and two adjacent vari-able domains (ETAS and CSB). All blocks of conserved sequences, ETAS 1, ETAS 2, CSB 1, CSB 2, and CSB 3 could be identified in Tupaia by aligning it to the respective mammalian consensus sequences proposed by Sbisa et al. (1997) (see also fig. 1).

We could define another 60-nt region 5′ of ETAS 1 (the L-strand is the reference) showing 66% similarity to the ETAS 1 region consensus of Sbisa et al. (1997) (indicated by a star in fig. 1). The same constellation can be observed in the rabbit mitochondrial genome. It remains to be tested in functional studies whether this sequence is of regulatory importance and whether its presence possibly reflects the evolutionary relatedness of rabbits and Tupaia. Otherwise, the high turnover rate in the ETAS domain with a frequent occurrence of de-letions and insertions precludes the use of the presence of this sequence in different taxa as a phylogenetic marker.

Adjacent to the CSB 1 block, a tandem repetitive sequence could be detected. Regarding the length of the repeated sequence, three equally good motifs could be detected in the Tupaia sequence. The 26-nt consensus

sequence is defined on the basis of eight different re-peats detected in the sequenced individual. The vari-ability between the single motifs can therefore be de-picted as YRCACAYAYRCRCAYRYRCRCAYAYR. It is apparent that parts of some of these motifs are similar to the repeat motifs of other taxa listed in Sbisa et al. (1997) and Pumo et al. (1998). Regarding the taxa in-cluded in our data set, tests carried out by us and others indicate that tandem repeats at the same location are absent only in the fin whale and all primates (Sbisa et al. 1997; Tagliaro et al. 1997). As for the ETAS domain, the high turnover rate of tandem repetitive DNA and the lack of a model of sequence evolution means that nei-ther the presence nor structural similarity of the tandem repeat observed when comparing different taxa can be taken as a sign of their phylogenetic relatedness.

It has to be kept in mind, though, that only one individual has been sequenced. Therefore, any conclu-sions about infraspecific length and repeat variability, possibly also detected in a heteroplasmic state in one individual, have to await further infraspecific studies. Since independent amplifications of target sequences with different lengths exhibited the same repeat se-quence, we have no indication that our PCR strategy possibly gives rise to artifactual sequences in the repeat array.

Phylogeny

Since our results argue against the classical hy-pothesis of a close phylogenetic relationship of Scan-dentia to primates, we took several precautions to avoid erratic results and false conclusions from our analyses. Our phylogenetic analyses of the interordinal mamma-lian relationships were based mainly on the 12 concat-enated aa sequences of the H-strand–encoded proteins to allow for a better statistical significance of the recon-struction and to avoid the varying topologies possibly generated by single mitochondrial gene analyses (Zar-doya and Meyer 1996; Cao et al. 1998). The L-strand–encoded ND6 gene, which exhibits significant devia-tions in its nucleotide and aa compositions, was exclud-ed from the analyses. It can therefore be assumed that all sequences under consideration basically follow the same model of sequence evolution.

The phylogenetic signal in the data set was tested a priori in a likelihood-mapping analysis revealing a low amount of phylogenetic noise in the data, with only 0.2 % of all quartets mapping into the region representing starlike evolution. In accordance with this, the topolo-gies (fig. 2) presented herein partly receive strong a pos-teriori support from the bootstrap and quartet puzzling values, although the internode distances of the deeper splits are small. Moreover, by grouping the taxa in dif-ferent clusters, thus asking for the support of the clade separating primates from Tupaia, the hypothesis of Scandentia being grouped together with primates re-ceives the least support from the data.

Different reconstruction algorithms (ML, MP, NJ) were applied for the same data sets. Additionally, we incorporated different models of sequence evolution im-

plemented in PUZZLE with and without assuming rate heterogeneity across sites for the ML reconstructions, as well as different distance corrections for the NJ analyses. The fact that all of these approaches, each of which relies on different assumptions, do not significantly alter the topology of the obtained trees shows that Tupaia is robustly positioned away from primates.

In all tree reconstructions, we identified the hedgehog as the basal relative to the other eutherians, as suggested by Mouchaty et al. (2000), followed by the mouse and the rat. To test whether the opossum sequence represents an adequate outgroup (see also Philippe 1997) and to determine how alternative outgroups affected the topology of our tree reconstruction, we used hedgehog, mouse, and rat sequences, alone and combined, as alternative outgroups. As a result, we could not observe an influence of alternative rootings on the resulting branching pattern.

Furthermore, we excluded the hedgehog, the rabbit, and the guinea pig from the analyses in order to sample only species which either fulfilled the aa frequency distribution as assumed in the ML model or were found to be robustly positioned in phylogenetic tree reconstructions based on complete mtDNA sequences. This resulted in topologies that were not different for the two alternatives; however, the removal of the three sequences from the data resulted in increased support for nodes of the tree. Due to the short time span of the major radiation of the recent eutherian orders, exclusion or inclusion of species can have serious effects on at least the reliability of the branching pattern (see also Arnason, Gullberg, and Janke 1999). This was also verified with a comparison of the ML tree (fig. 2A, left) and a user-defined tree grouping primates and Tupaia together in one clade. The user-defined tree was not significantly rejected in the 5% significance test of Kishino and Hasegawa (1989).

When the tree was pruned to only few representatives of the analyzed clades and the opossum was retained as the outgroup, a sister group relationship between primates and Tupaia was significantly rejected in both the ML significance test without rate heterogeneity categories (PUZZLE) and the MP tree length comparison (PAUP). However, inclusion of the rabbit sequence in the pruned data sets strongly affected the ML significance test. A Tupaia/primate sister group relationship was also significantly rejected with four and eight categories of rate heterogeneity.

Finally, phylogenetic analyses of the protein-coding nucleotide sequences corroborated the results of the aa reconstructions to a large degree. Although Tupaia, the armadillo, the rabbit, and the ferungulates/bat constitute one clade in this tree reconstruction, the exact relationship among them remains unresolved. Nevertheless, a close relationship of Tupaia and primates can be excluded (see fig. 2).

A test of the congruence of the results obtained from phylogenetic reconstructions based on two different mitochondrial data sets representing the polypeptides and the rRNA-specifying regions remains inconclusive. The likelihood-mapping analyses of the 12S rRNA and 16S rRNA genes alone and combined revealed only marginal phylogenetic signals. Splitting the data set into single- and double-stranded regions revealed the highest phylogenetic information content in the rRNA stem regions. The highest bootstrap reliability for this data subset was found for the primate monophyly (100%) and a sister group relationship of Tupaia and the rabbit, with up to 77% depending on the method of reconstruction. Further resolution could not be obtained with this data set, consistent with and extending the notion of McNiff and Allard (1998), who doubt the usefulness of 12S rDNA sequences in defining archontan relationships.

The major conclusion which can be drawn from our complete mtDNA analyses is that Tupaia and primates do not share a common ancestor exclusive of all other eutherians. Since the complete mtDNA essentially provides information of a single recombination unit, it is important to note that the interpretations obtained from the mtDNA data set receive support from molecular data already generated from different nuclear loci. Although the taxonomic sampling and the phylogenetic reconstruction methods were somewhat different from each other, data obtained from (1) exon 28 of the von Willebrand factor gene (Porter, Goodman, and Stanhope 1996), (2) the interphotoreceptor retinoid-binding protein (Stanhope et al. 1992), (3) eight polypeptide sequences (α-, β-, and embryonic α-hemoglobin chains, myoglobin, α-crystallin A chain, cytochrome c, pancreatic ribonuclease, and fibrinopeptide A or B; Czelusniak et. al. 1990), and (4) epsilon-globin (Bailey, Slightom, and Goodman 1992) rule out a sister group relationship between Scandentia and primates as well.

Molecular data apparently contradicting our hypothesis have been obtained from 208 and 266 sites of introns 4 and 5 of the Mhc DRB genes, respectively, in which a Scandentia/primate sister group relationship is weakly supported (Kupfermann et al. 1999). Incoming sequence data will help to determine whether this contradiction can possibly be explained by the peculiar mode of sequence evolution at these loci, giving rise to, for instance, a questionable orthology/paralogy of the sequences under consideration, or is due to the high stochastic fluctuations of the results obtained on the basis of short sequences with few informative sites.

Finally, discrepancies with our interpretation can be found in combined analyses of molecular and morphological data, as outlined by Liu and Miyamoto (1999), who evaluated the interordinal relationships of eutherian mammals. However, partitioning the data set and analyzing the morphological and molecular characters independently showed that the strong support for a Scandentia/primate sister group relationship was gained only from the morphological data (98% bootstrap score), whereas the molecular data are not able to resolve the respective relationships in an MP analysis.

Overall, our data add to the growing evidence that the acceptance of Tupaia as an appropriate outgroup for primates needs to be reconsidered. Therefore, the question arises as to which of the investigated eutherian orders is the most likely sister group to Scandentia. The newly established complete mtDNA data support the

close relationship between Lagomorpha and Scandentia, which was tentatively proposed by Graur, Duret, and Gouy (1996). Although this conclusion needs to be discussed in light of insufficient taxonomic sampling in the four-taxon approach, considerable branch length differences, and influences of missing and uninformative data (Halanych 1998), other reports, based on single nuclear loci, add further support to this hypothesis (Porter, Goodman, and Stanhope 1996; Bailey, Slightom, and Goodman 1992; Czelusniak et al. 1990). Supposing that a closer relationship of Scandentia to Lagomorpha than to primates reflects the true evolutionary history, the original phylogenetic classification proposing an archontan monophyly is rendered invalid from the complete mtDNA analyses of both bats and tree shrews.

Acknowledgments

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. **42**:459–468.

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.

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.

———. 1999. The mitochondrial DNA molecule of the aardvark, *Orycteropus afer,* and the position of the Tubulidentata in the eutherian tree. Proc. R. Soc. Lond. B Biol. Sci. **266**:339–345.

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 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.

BAILEY, W. J., J. L. SLIGHTOM, and M. GOODMAN. 1992. Rejection of the ''flying primate'' hypothesis by phylogenetic evidence from the ε-globin gene. Science **256**:86–89.

BENSON, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27**:573–580.

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.

CAO, Y., A. JANKE, P. J. WADDELL, M. WESTERMAN, O. TAKENAKA, S. MURATA, N. OKADA, S. PÄÄBO, and M. HASEGAWA. 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. J. Mol. Evol. **47**:307–322.

CZELUSNIAK, J., M. GOODMAN, B. F. KOOP, D. A. TAGLE, J. SHOSHANI, G. BRAUNITZER, T. K. KLEINSCHMIDT, W. W. DE JONG, and G. MATSUDA. 1990. Perspectives from amino acid and nucleotide sequences on cladistic relationships among higher taxa of Eutherian. Pp. 545–572 *in* H. H. GENOWAYS, ed. Current mammalogy. Plenum Press, New York.

D'ERCHIA, A. M., C. GISSI, G. PESOLE, C. SACCONE, and U. ARNASON. 1996. The guinea-pig is not a rodent. Nature **381**:597–599.

FEARNLEY, I. M., and J. E. WALKER. 1987. Initiation codons in mammalian mitochondria: differences in genetic code in the organelle. Biochemistry **26**:8247–8251.

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

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.

GISSI, C., A. GULLBERG, and U. ARNASON. 1998. The complete mitochondrial DNA sequence of the rabbit *Oryctolagus cuniculus.* Genomics **50**:161–169.

GRAUR, D., L. DURET, and M. GOUY. 1996. Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). Nature **379**:333–335.

GUTELL, R. R., and G. E. FOX. 1988. A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res. **16**:175–313.

HALANYCH, K. M. 1998. Lagomorphs misplaced by more characters and fewer taxa. Syst. Biol. **47**:138–146.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

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

KAY, R. F., R. W. THORINGTON, and P. HOUDE. 1990. Eocene plesiadapiform shows affinities with flying lemurs not primates. Nature **345**:342–344.

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.

KRETTEK, A., A. GULLBERG, and U. ARNASON. 1995. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus,* and the phylogenetic position of Lipotyphla. J. Mol. Evol. **41**:952–957.

KUPFERMANN, H., Y. SATTA, N. TAKAHATA, H. TICHY, and J. KLEIN. 1999. Evolution of Mhc-DRB introns: implications for the origin of primates. J. Mol. Evol. **48**:663–674.

LIU, F.-G., and M. MIYAMOTO. 1999. Phylogenetic assessment of molecular and morphological data for eutherian mammals. Syst. Biol. **48**:54–64.

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

MCNIFF, B. E., and M. W. ALLARD. 1998. A test of archonta monophyly and the phylogenetic utility of the mitochondrial gene 12S rRNA. Am. J. Phys. Anthropol. **107**:225–241.

MARTIN, R. D. 1990. Primate origin and evolution: a phylogenetic reconstruction. Chapman Hall, London.

MOUCHATY, S. K., A. GULLBERG, A. JANKE, and U. ARNASON. 2000. The phylogenetic position of the Talpidae within eutheria based on analysis of complete mitochondrial sequences. Mol. Biol. Evol. **17**:60–67.

NOVACEK, M. J. 1992. Mammalian phylogeny: shaking the tree. Nature **356**:121–125.

OJALA, D., J. MONTOYA, and G. ATTARDI. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature **290**:470–474.

PHILIPPE, H. 1997. Rodent monophyly: pitfalls of molecular phylogenies. J. Mol. Evol. **45**:712–715.

PORTER, A. P., M. GOODMAN, and M. J. STANHOPE. 1996. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gen. Mol. Phylogenet. Evol. **5**:89–101.

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.

SACCONE, C., C. LANAVE, G. PESOLE, and G. PREPARATA. 1990. Influence of base composition on quantitative estimates of gene evolution. Methods Enzymol. **183**:570–583.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

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.

SPRINGER, M. S., and E. DOUZERY. 1996. Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. J. Mol. Evol. **43**:357–373.

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.

STANHOPE, M. J., J. CZELUSNIAK, J.-S. SI, J. NICKERSON, and M. GOODMAN. 1992. A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. Mol. Phylogenet. Evol. **1**:148–160.

STARCK, D. 1978. Vergleichende Anatomie der Wirbeltiere auf evolutionsbiologischer Grundlage. Springer-Verlag, Berlin and New York.

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

———. 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence. Proc. Natl. Acad. Sci. USA **94**:6815–6819.

SWOFFORD, D. L. 1999. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

TAGLIARO, C. H., M. P. SCHNEIDER, H. SCHNEIDER, I. C. SAMPAIO, and M. J. STANHOPE. 1997. Marmoset phylogenetics, conservation perspectives, and evolution of the mtDNA control region. Mol. Biol. Evol. **14**:674–684.

TEELING, E. C., M. SCALLY, D. J. KAO, M. L. ROMAGNOLI, M. S. SPRINGER, and M. J. STANHOPE. 2000. Molecular evidence regarding the origin of echolocation and flight in bats. Nature **403**:188–192.

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.

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.

ZARDOYA, R., and A. MEYER. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. **13**:933–942.

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

ROSS CROZIER, reviewing editor