

Mitochondrial Versus Nuclear Gene Sequences in Deep-Level Mammalian Phylogeny Reconstruction

Mark S. Springer,* Ronald W. DeBry,† Christophe Douady,‡ Heather M. Amrine,* Ole Madsen,§ Wilfried W. de Jong,§|| and Michael J. Stanhope‡¶

*Department of Biology, University of California at Riverside; †Department of Biological Sciences, University of Cincinnati; ‡Queen's University of Belfast, Biology and Biochemistry, Belfast, Ireland; §Department of Biochemistry, University of Nijmegen, the Netherlands; ||Institute for Systematics and Population Biology, Amsterdam, the Netherlands; ¶Bioinformatics, SmithKline Beecham Pharmaceuticals, Pennsylvania

Both mitochondrial and nuclear gene sequences have been employed in efforts to reconstruct deep-level phylogenetic relationships. A fundamental question in molecular systematics concerns the efficacy of different types of sequences in recovering clades at different taxonomic levels. We compared the performance of four mitochondrial data sets (cytochrome *b*, cytochrome oxidase II, NADH dehydrogenase subunit I, 12S rRNA–tRNA^{Val}–16S rRNA) and eight nuclear data sets (exonic regions of α -2B adrenergic receptor, aquaporin, β -casein, γ -fibrinogen, interphotoreceptor retinoid binding protein, κ -casein, protamine, von Willebrand Factor) in recovering deep-level mammalian clades. We employed parsimony and minimum-evolution with a variety of distance corrections for superimposed substitutions. In 32 different pairwise comparisons between these mitochondrial and nuclear data sets, we used the maximum set of overlapping taxa. In each case, the variable-length bootstrap was used to resample at the size of the smaller data set. The nuclear exons consistently performed better than mitochondrial protein and rRNA–tRNA coding genes on a per-residue basis in recovering benchmark clades. We also concatenated nuclear genes for overlapping taxa and made comparisons with concatenated mitochondrial protein-coding genes from complete mitochondrial genomes. The variable-length bootstrap was used to score the recovery of benchmark clades as a function of the number of resampled base pairs. In every case, the nuclear concatenations were more efficient than the mitochondrial concatenations in recovering benchmark clades. Among genes included in our study, the nuclear genes were much less affected by superimposed substitutions. Nuclear genes having appropriate rates of substitution should receive strong consideration in efforts to reconstruct deep-level phylogenetic relationships.

Introduction

Higher-level mammalian phylogenetics remains an outstanding problem in systematics. Anatomical data suggest a moderately well resolved tree that includes several supraordinal groupings for the 18 orders of living placental mammals (Novacek 1992; Shoshani and McKenna 1998). Molecular sequence data have corroborated clades such as Cetartiodactyla (de Jong 1998; Gatesy et al. 1999; Liu and Miyamoto 1999) and Paenungulata (Lavergne et al. 1996; Springer et al. 1997a, 1997b), challenged other morphological hypotheses (Graur, Hide, and Li 1991; D'Erchia et al. 1996; Arnason, Gullberg, and Janke 1997; Janke, Xu, and Arnason 1997; Springer et al. 1997a, 1997b; Stanhope et al. 1998b; Mouchaty et al. 2000; Teeling et al. 2000), and suggested new supraordinal associations (Graur and Higgins 1994; Porter, Goodman, and Stanhope 1996; Arnason, Gullberg, and Janke 1997; Springer et al. 1997a, 1997b; Stanhope et al. 1998b; Liu and Miyamoto 1999; Waddell, Okada, and Hasegawa 1999; Mouchaty et al. 2000). Recent attention in higher level mammalian and vertebrate molecular phylogenetics has focused on either protein-coding sequences from complete mitochondrial (mt) genomes (e.g., Krettek, Gullberg, and Arnason 1995; D'Erchia et al. 1996; Zardoya and Meyer 1997; Rasmussen, Janke, and Arnason 1998; Arnason,

Gullberg, and Janke 1999), mt rRNA gene sequences (Lavergne et al. 1996; Springer et al. 1997a; Stanhope et al. 1998a, 1998b) or exonic regions of nuclear (nc) protein-coding genes (Porter, Goodman, and Stanhope 1996; Madsen et al. 1997; Stanhope et al. 1996, 1998a, 1998b; Springer et al. 1997a, 1997b; Gatesy et al. 1999). A fundamental question that remains in this field is whether different types of mt or nc sequences are more reliable for recovering deep-level relationships such as those among orders of eutherian mammals (Springer et al. 1999).

Molecular sequence data can be categorized in different ways, including by function (protein-coding vs. noncoding vs. structural RNA) and by genome (e.g., mitochondrial vs. nuclear). Different categories of sequences have distinct properties. In many groups, including mammals, the rate of nucleotide substitution among mt protein-coding genes is generally more rapid than the rate of nucleotide substitution among protein-coding regions of nc genes (Vawter and Brown 1986). A second difference is that the mt genome is inherited as a single, haploid, nonrecombining linkage unit. This haploid mode of inheritance leads to a smaller effective population size compared with the nc genome. A smaller effective population size, in turn, results in a shorter expected coalescence time for mt loci compared with nc loci and a greater probability that the mt gene tree will accurately reflect the species tree for closely spaced internodes compared with nc gene trees (Moore 1995). Arnason, Gullberg, and Janke (1999) made specific reference to the problem of inferring the deep relationships among mammalian orders and suggested that mt data

Key words: Mammalia, mitochondrial DNA, nuclear DNA, phylogeny reconstruction.

Address for correspondence and reprints: Mark Springer, Department of Biology, University of California, Riverside, California 92521. E-mail: mark.springer@ucr.edu.

Mol. Biol. Evol. 18(2):132–143, 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

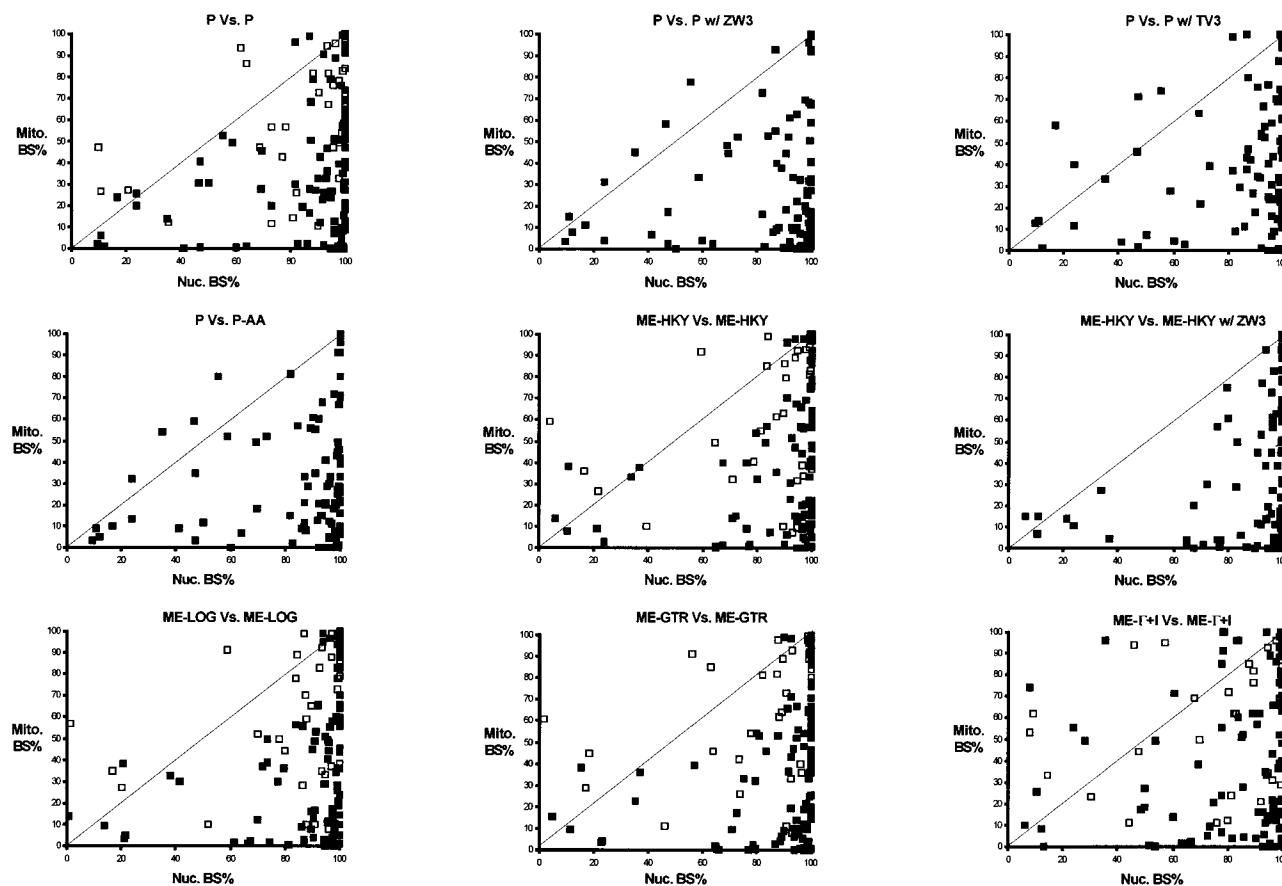


FIG. 1.—Plots of mitochondrial bootstrap support (Mito. BS%) versus nuclear bootstrap support (Nuc. BS%) for benchmark clades for nine different comparisons. Data are from table 1. In the title for each panel, the method for the nuclear sequences is listed first, and the method for the mitochondrial sequences is listed second. Filled squares, which occur in all nine plots, are for mitochondrial protein-coding genes versus exons of nuclear genes. Open squares, which occur in five of nine plots, are for mitochondrial RNA genes versus exons of nuclear genes. AA = amino acids; GTR = maximum-likelihood distances under the general time reversible model of sequence evolution; HKY = maximum-likelihood distances under the HKY85 model of sequence evolution; LOG = logdet distances; ME = minimum evolution; P = parsimony; TV3 = transversion parsimony at third codon positions; ZW3 = zero weight at third codon positions; $\Gamma+I$ = maximum-likelihood distances under the HKY85 model of sequence evolution with an allowance for invariant sites and a gamma distribution of rates. A table that gives bootstrap support values for each benchmark clade for each phylogenetic method for all pairwise comparisons is available from M.S.S.

Table 1
Summary of Comparisons Between Mitochondrial and Nuclear Genes

Comparison	Nc > Mt	Nc = Mt	Nc < Mt	Mean Nc Bootstrap %	Mean MT Bootstrap %	Difference
Nc vs. Mt protein-coding						
P vs. P	100	1	5	86.7	36.3	50.4
P vs. P with ZW3	99	1	6	86.7	30.7	56.0
P vs. P with TV3	96	1	9	86.7	43.4	43.3
P vs. P-AA	100	2	4	86.7	34.7	51.9
ME-HKY vs. ME-HKY	98	2	6	89.6	42.6	47.0
ME-HKY vs. ME-HKY with ZW3	99	5	2	89.6	35.3	54.3
ME-LOG vs. ME-LOG	97	3	6	89.5	42.0	47.5
ME-GTR vs. ME-GTR	99	3	4	89.4	41.3	48.1
ME- $\Gamma+I$ vs. ME- $\Gamma+I$	90	1	15	84.3	43.8	40.5
Nc vs. Mt RNA						
P vs. P	33	3	6	84.7	60.4	24.3
ME-HKY vs. ME-HKY	32	3	7	85.9	67.6	18.3
ME-LOG vs. ME-LOG	32	3	7	86.2	66.5	19.7
ME-GTR vs. ME-GTR	31	4	7	84.7	67.5	17.2
ME- $\Gamma+I$ vs. ME- $\Gamma+I$	32	3	7	80.9	64.4	16.5

NOTE.—Nc > Mt, number of cases for which nuclear genes provided higher bootstrap support than mitochondrial genes for a benchmark clade; Nc = Mt, number of cases for which nuclear genes and mitochondrial genes provided equivalent bootstrap support for a benchmark clade; Nc < Mt, number of cases for which nuclear genes provided lower bootstrap support than mitochondrial genes for a benchmark clade. Abbreviations for methods are given in the legend to figure 1.

A

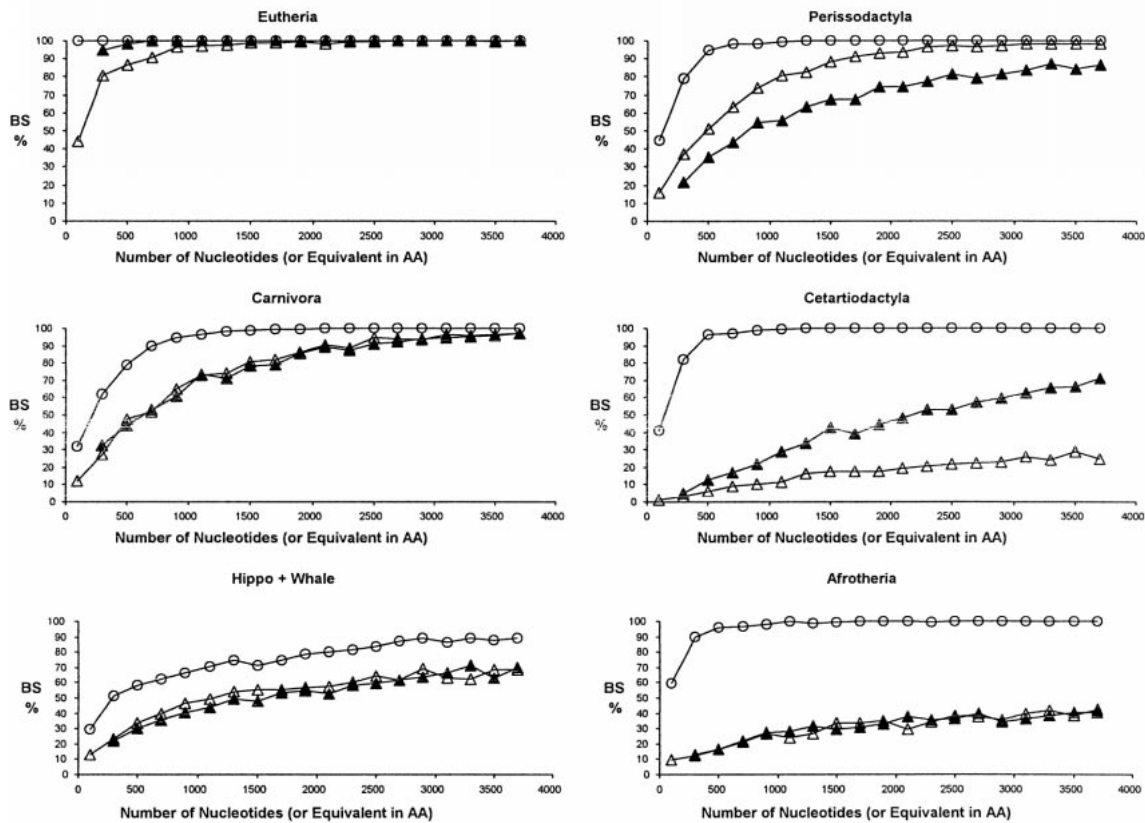


FIG. 2.—Plots of bootstrap support percentages (BS%) as a function of the number of resampled nucleotides (or equivalent in amino acids) for four benchmark clades and two more controversial hypotheses (hippo + whale; Afrotheria). Analyses were based on concatenated mitochondrial and nuclear sequences, respectively, for 20 overlapping taxa (see *Materials and Methods*). Results for nuclear genes are shown in open circles; results for mitochondrial genes are shown in triangles. (A) Parsimony and (B) minimum evolution with maximum-likelihood distances under the HKY85 model of sequence evolution. In A, open triangles are for nucleotide parsimony, and filled triangles are for amino acid parsimony. In B, open triangles are for equal weights at all codon positions, and filled triangles are for zero weight at third codon positions.

would be more efficient; i.e., nc data sets must be larger than mt data sets to achieve commensurate levels of resolution.

Several studies have compared the relative performances of individual mt genes in higher level phylogenetics using metrics such as the bootstrap support percentage and retention indices (Cao et al. 1994; Cummings, Otto, and Wakeley 1995; Zardoya and Meyer 1996; Naylor and Brown 1997). In this paper, we take an empirical approach to investigate the performance of several well-sampled mt and nc genes in recovering well-supported, deep-level mammalian clades. We find that the nc exons are more efficient and consistently achieve greater resolving power on a per-residue basis in comparison with equivalent lengths of either mt protein-coding genes or mt rRNA genes.

Materials and Methods

12S rRNA, tRNA^{Val}, and 16S rRNA sequences were obtained for *Dobsonia moluccensis* (16S rRNA

only; AF179290) and *Tonatia bidens* (AF179288) as described elsewhere (Springer, Hollar, and Burke 1995; Springer et al. 1997b). New interphotoreceptor retinoid binding protein (IRBP) exon 1 sequences were obtained for *Erethizon dorsatum* (AF179292), *Tapirus pinchaque* (AF179294), and *Vulpes velox* (AF179293) following Stanhope et al. (1992, 1996). Additional sequences for these and other genes were extracted from GenBank. The additional mt genes included cytochrome *b*, cytochrome oxidase subunit II (COII), and NADH dehydrogenase subunit I (ND1); these mt protein-coding genes were chosen because of their extensive taxonomic representation. Zardoya and Meyer (1996) classified mt protein-coding genes into three groups (good, medium, and poor) based on their performance in recovering phylogenetically distant relatives. The mt protein-coding genes incorporated in our analysis included one gene (cytochrome *b*) in the good group and two genes (COII, ND1) in the medium group. Additional nc gene sequences included protein-coding regions of α -2B adrenergic receptor (A2AB), aquaporin, β -casein, γ -fibrino-

B

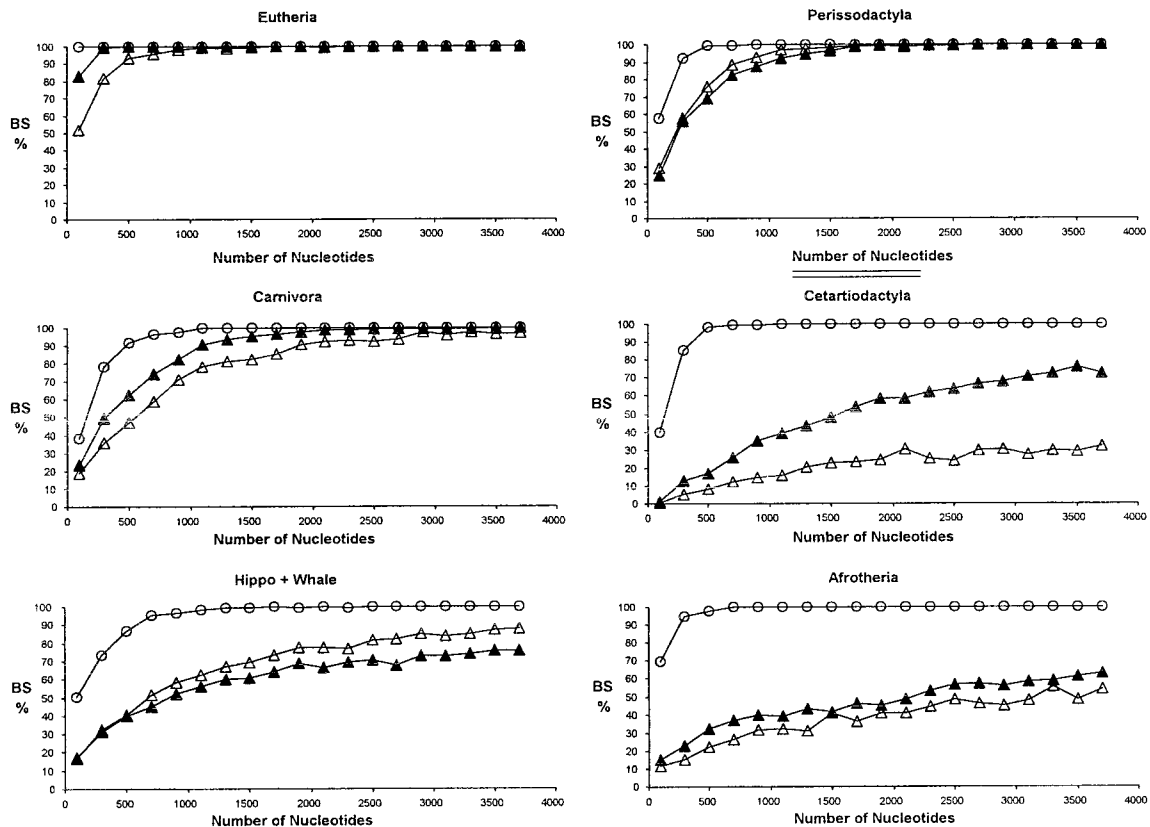


FIG. 2 (Continued)

gen, κ -casein, protamine, and von Willebrand Factor (vWF). Among the nuclear genes, A2AB, IRBP, and vWF included representatives of all placental orders and marsupial outgroups. CLUSTAL W (Thompson, Higgins, and Gibson 1994) was used to align 12S rRNA-tRNA^{Val}-16S rRNA, COII, cytochrome *b*, ND1, A2AB, IRBP, and vWF sequences. Ambiguous regions of the 12S rRNA-tRNA^{Val}-16S rRNA and COII alignments were excluded from the analyses. A short glutamic acid repeat region of the A2AB gene was also difficult to align and was omitted following Stanhope et al. (1998a). The aquaporin data set was from Madsen et al. (1997) and Stanhope et al. (1998a), with additional sequences from GenBank (AJ277647, AJ251100, AJ251101). The β -casein, γ -fibrinogen, κ -casein, and protamine data sets were from WHIPPO-2 (Gatesy et al. 1999), with ambiguous regions of the alignment omitted following Gatesy et al. (1999). After removal of the ambiguous regions, the aligned data sets were the following lengths: 12S rRNA-tRNA^{Val}-16S rRNA, 2,006 bp; cytochrome *b*, 1,152 bp; COII, 675 bp; ND1, 957 bp; A2AB, 1,164 bp; aquaporin, 333 bp; β -casein, 520 bp; γ -fibrinogen, 679 bp; IRBP, 1,292 bp; κ -casein, 519 bp; protamine, 386 bp; and vWF, 1,251 bp. Alignments, as well as ac-

cession numbers for all sequences, are available on request from mark.springer@ucr.edu.

Phylogenetic analyses were performed for 32 different pairwise comparisons between an mt data set and an nc data set. In each case, we used the maximum set of overlapping eutherian and metatherian taxa (appendix); some taxa were chimeric (appendix). For example, *Elephas* and *Loxodonta* are both in the order Proboscidea, and it was necessary to use a nuclear sequence for *Elephas* and a mitochondrial sequence for *Loxodonta* in some comparisons. For all pairwise comparisons, the variable-length bootstrap was used to resample at the size of the smaller data set. For example, in the comparison between A2AB and COII, we resampled both data sets at the size of COII because it was the smaller (675 bp) of the two data sets. Use of the variable-length bootstrap permitted us to assess performance on a per-residue basis. We evaluated the performance of the mt and the nc data sets, respectively, in recovering deep-level, benchmark clades where monophyly is generally agreed upon and where available sequences index intra-clade divergences that are no later than the Eocene based on the current fossil record (McKenna and Bell 1997). One exception is the Australasian marsupial order Di-

Table 2
Uncorrected Percentages of Sequence Divergence for
Interordinal Comparisons

Gene	Pos/Sub	Bos to Balaen- optera	Cani- form to Feli- form	Dugong to Proca- via	Placen- tals to Didel- phis	Δ	
Mitochondrial genes							
COII	All/all	17.6	19.7	NA	24.0	5.4	
	1+2/TV	1.3	0.9	NA	4.4	3.3	
	1+2/TS	4.2	2.9	NA	7.0	3.5	
	3/TV	18.9	19.8	NA	24.3	5.0	
CYTB ..	3/TS	22.9	31.7	NA	25.1	-2.2	
	All/all	18.7	20.3	20.8	25.3	5.4	
	1+2/TV	2.2	2.2	4.2	5.9	3.0	
	1+2/TS	4.9	5.8	5.1	6.9	1.6	
ND1	3/TV	20.7	19.8	24.6	26.4	4.7	
	3/TS	21.3	25.2	19.5	24.1	2.1	
	All/all	19.0	18.2	NA	27.4	8.8	
	1+2/TV	2.5	1.6	NA	7.2	5.2	
RNA	1+2/TS	6.1	3.7	NA	7.8	2.9	
	3/TV	18.2	18.8	NA	30.9	12.4	
	3/TS	21.7	25.1	NA	21.0	-2.4	
	All/all	9.8	7.8	9.2	15.8	6.9	
Nuclear genes	All/TV	2.9	2.4	2.7	6.9	4.2	
	All/TS	6.9	5.4	6.5	8.9	2.6	
	A2AB ...	All/all	7.1	6.8	9.0	26.3	18.7
	1+2/TV	1.9	1.8	2.5	8.8	6.7	
IRBP ...	1+2/TS	1.4	1.4	3.0	6.7	4.8	
	3/TV	6.4	5.5	4.7	24.4	18.9	
	3/TS	8.3	8.6	11.3	23.3	14.0	
	All/all	11.9	11.9	12.9	29.1	16.9	
vWF	1+2/TV	2.1	1.5	2.5	9.1	7.1	
	1+2/TS	4.6	4.3	3.9	10.3	6.0	
	3/TV	6.3	10.8	10.2	23.4	14.3	
	3/TS	15.9	13.2	15.6	25.1	10.2	
vWF	All/all	11.9	11.3	12.4	28.3	16.4	
	1+2/TV	2.0	1.9	3.2	8.8	6.4	
	1+2/TS	4.3	3.8	4.8	10.7	6.4	
	3/TV	8.4	6.2	5.4	19.1	12.4	
vWF	3/TS	14.9	16.5	15.7	26.7	11.0	

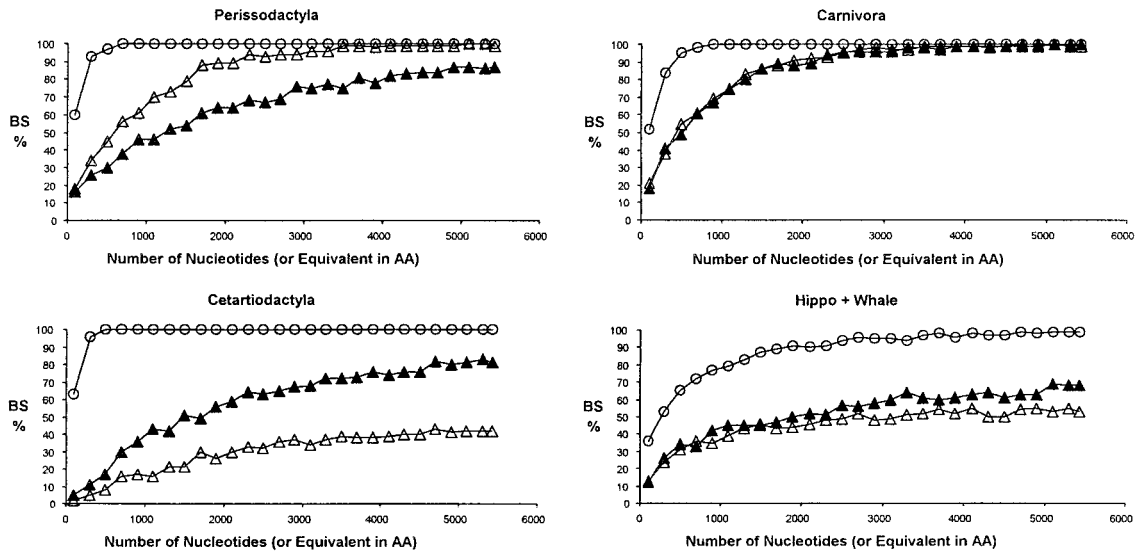
NOTE.—NA = not available; Pos = positions; Sub = substitutions (TV = transversions; TS = transitions); Δ = placental-to-*Didelphis* divergence. To standardize the comparisons, all calculations were performed after excluding positions in the alignment with missing/ambiguous data for any of the seven taxa (A2AB, 1,088 bp; IRBP, 1,004 bp; vWF, 1,114 bp; RNA, 1,979 bp; COII, 681 bp; CYTB, 1,005 bp; ND1, 942 bp).

protodontia, for which intraclade divergences based on first fossil occurrences are Oligocene (Woodburne et al. 1993; Woodburne and Case 1996). However, most workers accept an earlier age based on the incompleteness of the Paleogene Australian fossil record coupled with molecular-clock estimates (Springer and Kirsch 1991; Woodburne and Case 1996; Kirsch, Lapointe, and Springer 1997); the latter consistently place the split between vombatiform and phalangeriform diprotodontians in the late Paleocene or Eocene (Springer and Kirsch 1991; Kirsch, Lapointe, and Springer 1997). Benchmark clades that were evaluated were as follows: Carnivora (canoids + feloids), Cetacea (mysticetes + odontocetes), Cetartiodactyla (artiodactyls + cetaceans), Chiroptera (megachiropterans + microchiropterans), Diprotodontia (phalangeriforms + vombatiforms), Eutheria (a minimum of 13 of 18 placental orders were represented in tests to recover Eutheria), Paenungulata (hyracoids +

proboscideans + sirenians), Perissodactyla (ceratomorphs + hippomorphs), Primates (strepsirhines + haplorhines), Ruminantia (tragulines + pecorans), Suina (suids + tayassuids), and Xenarthra (infraorder Pilosa + infraorder Cingulata; *sensu* Nowak and Paradiso 1983). Chiropteran monophyly and the paenungulate hypothesis have previously been questioned (e.g., Petti-grew [1986] challenged bat monophyly, and Prothero and Schoch [1989] argued against the Paenungulata), but recent work provides compelling support in favor of these clades (Wible and Novacek 1988; Stanhope et al. 1992, 1996, 1998a, 1998b; Simmons 1994; Kirsch et al. 1995; Lavergne et al. 1996; Porter, Goodman, and Stanhope 1996; Springer et al. 1997b, 1999). Afrotheria, Eulipotyphla, hippo + Cetacea, and Rodentia were not scored as benchmark clades because of controversy surrounding these hypotheses. Afrotheria is supported by molecular data (Stanhope et al. 1998b), but not by morphological data (Asher 1999); Eulipotyphla (Erinaceidae + Soricidae + Talpidae + Solenodontidae; Waddell, Okada, and Hasegawa 1999) is supported by nuclear genes (Stanhope et al. 1998b), but not by mt genome data (Mouchaty et al. 2000); hippo + Cetacea is supported by molecular data, but not by morphological data (O'Leary and Geisler 1999); and rodent monophyly remains controversial (Graur, Hide, and Li 1991; D'Erchia et al. 1996; Reyes, Pesole, and Saccone 1998; Penny et al. 1999).

In addition to pairwise comparisons between individual mt and nc data sets, we compared concatenated nc genes with concatenated mt protein-coding genes from mt genomes. First, concatenated sequences for the three nc genes with the most extensive taxonomic representation (A2AB, IRBP, vWF; total length = 3,707 bp) were compared with mt protein-coding genes (total length = 9,828 bp); this comparison included 20 overlapping taxa and permitted scoring of support for four benchmark clades (Carnivora, Cetartiodactyla, Eutheria, and Perissodactyla). Overlapping taxa were as follows: Bos, Canioidea (Canis, Phoca, or Vulpes), Ceratomorpha (Ceratotherium, Dicerorhinus, or Tapirus), Didelphis, Diprotodontia (Macropus or Vombatus), Equus, Erinaceus, Felis, Hippopotamus, Homo, Hystricognathi (Cavia, Dasyprocta, or Erethizon), Muridae (Mus or Rattus), Mysticeti (Balaenoptera or Megaptera), Orycteropus, Oryctolagus, Phyllostomidae (Artibeus, Macrotus, or Tonatia), Proboscidea (Elephas or Loxodonta), Soricomorpha (Sorex, Scalopus, or Talpa), Sus, and Xenarthra (Bradypus or Dasypus). We also examined support for the controversial hypotheses Afrotheria and hippo + Cetacea. Second, we added β -casein, γ -fibrinogen, and κ -casein to the nuclear concatenation (total length = 5,425 bp). This resulted in 11 overlapping taxa between the mt and nc concatenations as follows: Bos, Canioidea (Canis, Phoca, or Vulpes), Ceratomorpha (Ceratotherium, Dicerorhinus, Dicerorhinus, or Tapirus), Equus, Felis, Hippopotamidae (Hexaprotodon or Hippopotamus), Homo, Muridae (Mus or Rattus), Mysticeti (Balaenoptera or Megaptera), Orycteropus, and Sus. With these 11 taxa, it was possible to score support for the benchmark clades Carnivora, Cetartiodactyla, and Perissodactyla, as

A



B

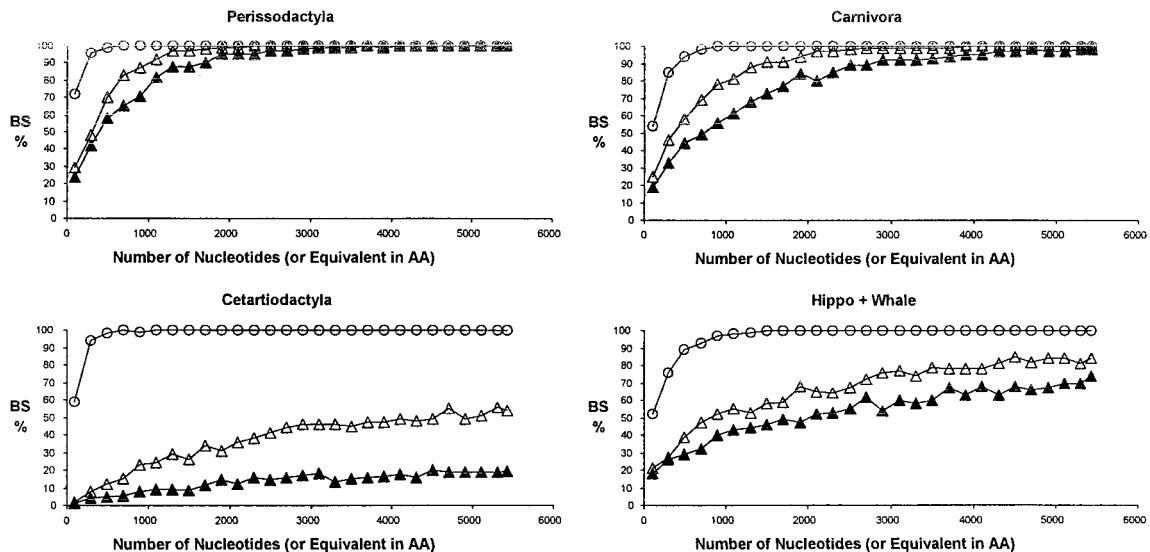


FIG. 3.—Plots of bootstrap support percentages (BS%) as a function of the number of resampled nucleotides (or equivalent in amino acids) for three benchmark clades and the more controversial hippo + whale hypothesis. Analyses were based on concatenated mitochondrial and nuclear sequences, respectively, for 11 overlapping taxa (see *Materials and Methods*). Results for nuclear genes are shown in open circles; results for mitochondrial genes are shown in triangles. (A) Parsimony and (B) minimum evolution with maximum-likelihood distances under the HKY85 model of sequence evolution. In A, open triangles are for nucleotide parsimony, and filled triangles are for amino acid parsimony. In B, open triangles are for equal weights at all codon positions, and filled triangles are for zero weight at third codon positions.

well as the more controversial hippo + Cetacea hypothesis. To obtain a concatenation of mt protein-coding sequences, we used the Mouchaty et al. (2000) mt data set and added *Loxodonta* (African elephant; AJ224821). Ambiguous and gapped regions were not included in the analyses following Mouchaty et al. (2000).

Phylogenetic analyses included unweighted and weighted parsimony and minimum evolution with logdet

(Lockhart et al. 1994) and maximum-likelihood distances, the latter under both the HKY85 (Hasegawa, Kishino, and Yano 1985) and the GTR models of sequence evolution with maximum-likelihood estimates of the substitution parameterizations. The HKY85 model was also employed with additional allowances for a gamma distribution (Γ) of rates and a fraction of invariant sites (I); both of these parameters were estimated with max-

imum likelihood from minimum-length parsimony trees. Third codon positions of mt protein-coding genes are known to be saturated, or nearly so, at deep phylogenetic levels (see below). We therefore examined the effect of downweighting third positions by either assigning zero weight or employing transversion parsimony. We also employed amino acid parsimony with the mt protein-coding genes. In all analyses, branches were swapped using the tree bisection-reconnection branch-swapping option. Parsimony analyses employed 10 random input orders per replicate, with gaps treated as missing data. All phylogenetic analyses were performed with PAUP, versions 4.0b2 and 4.0b3 (Swofford 1998).

For data sets that included marsupial outgroups, we compared the percentages of sequence divergence for three intraordinal/interordinal splits within Eutheria, all of which are in the range of 50–60 Myr based on the fossil record and molecular data (Bos to Cetacea [Balaeoptera]; Felis to Canis; Hyracoidea [Procavia] to Sirenia [Dugong]) (Garland et al. 1993; Arnason and Gullberg 1996; Springer 1997), with the divergence between these same placentals and a marsupial (*Didelphis*). The timing of the latter split is 97–160 Myr based on different types of data, with most estimates in the range of 100–130 Myr (Rowe 1993; Springer 1997; Kumar and Hedges 1998), approximately double that for the three intraplacental splits. Δ values were calculated as the increase in sequence divergence (uncorrected) in the placental-marsupial comparison relative to the mean value for the three intraplacental comparisons. Δ values index the amount of sequence divergence that is available for resolving relationships that are deeper than the three intraplacental splits but shallower than the marsupial-placental split. Δ values were calculated with uncorrected distances to highlight the effects of superimposed substitutions. Corrected distances, by definition, are employed to eliminate the effects of superimposed substitutions.

Results

Figure 1 and table 1 show the results of comparisons between nc and mt genes with different methods of phylogenetic analysis. Figure 1 plots bootstrap support values for nc versus mt genes for all of the phylogenetic methods that we employed. Nine of nine panels show comparisons between nc genes and mt protein-coding genes (filled squares); five of nine panels also show comparisons between nc genes and mt RNA genes (open squares). Data points that fall on the diagonal lines correspond to cases where nc and mt bootstrap support percentages are equivalent. Data points above the diagonal lines show cases where mt support exceeds nc support. Data points below the diagonal lines show cases where nc support exceeds mt support. Table 1 summarizes the numbers of cases for which nc bootstrap support was higher for a benchmark clade, the numbers of cases for which bootstrap support was equal, and the numbers of cases for which mt bootstrap support exceeded nc support. Table 1 also shows mean bootstrap

support (for all benchmark clades) for both nc and mt genes with different phylogenetic methods.

For comparisons between mt protein-coding genes and nc genes, there were 106 benchmark clades that were scored across 24 pairwise comparisons between individual nc and mt data sets. Support deriving from the nc genes exceeded support deriving from the mt protein-coding genes in 90–100 of the 106 cases with different phylogenetic methods (fig. 1 and table 1). A dense concentration of data points (filled squares) below the diagonal lines is evident in all of the panels in figure 1 and illustrates the superior performance of the nc genes. Mean bootstrap support was always higher for the nc genes (table 1). Whereas mean mt support ranged from 30.7% (parsimony with zero weight at third codon positions) to 43.8% (minimum evolution- $\Gamma+I$), nc support ranged from 84.3% (minimum evolution- $\Gamma+I$) to 89.6% (minimum evolution-HKY85). The difference in mean bootstrap ranged from 40.5% (minimum evolution- $\Gamma+I$ vs. minimum evolution- $\Gamma+I$) to 56.0% (parsimony vs. parsimony with zero weight at third codon positions).

Mitochondrial RNA genes performed better than mt protein-coding genes in comparisons against the nc genes. Benchmark clades were scored across eight pairwise comparisons between individual nc and mt RNA data sets. Bootstrap support scores were higher for nc genes in 31–33 of the 42 cases (fig. 1 and table 1). Mean support scores for nc genes were higher than those for mt RNA genes with all phylogenetic methods. Nevertheless, differences between means were always less than those for comparisons between nc genes and mt protein-coding genes. Differences in mean support ranged from 16.5% to 24.3% (table 1).

In all comparisons between the 3 or 6 concatenated nc genes and the 12 concatenated mt protein-coding genes from complete mt genomes, both nc and mt genes showed increased support for all of the benchmark clades as a function of the number of resampled nucleotides. However, the nc genes were always more efficient than the mt genes in recovering benchmark clades (figs. 2 and 3). This was especially noticeable for *Cetartiodactyla*, with both parsimony (figs. 2A and 3A) and minimum evolution (figs. 2B and 3B). The more controversial clades (*Afrotheria*; *Hippopotamidae* + *Cetacea*) were also recovered with greater efficiency with the nc data (figs. 2 and 3).

Among the data sets that included marsupial outgroups, Δ values for all substitutions were highest for the three nc genes. Furthermore, these values exceeded sequence divergence values for the intraplacental comparisons for all three nc genes (table 2). Among the mt protein-coding genes, Δ values (all substitutions) were always less than the intraplacental divergences (table 2).

Among nc genes, transitions and transversions showed positive Δ values both at first + second and at third positions. For transversions, Δ values both at first + second and at third positions were higher than intraplacental divergences. For transitions, Δ values at first + second positions were always higher than intraplacental divergences; at third positions, transition Δ values were higher than intraplacental divergences for A2AB

but lower than intraplacental divergences for IRBP and vWF.

Among mt protein-coding genes, transversion Δ values were higher than intraplacental divergences at first + second positions but lower than intraplacental divergences at third positions. Transition Δ values at first + second positions were lower than intraplacental divergence values except in one instance—the caniform-to-feliform divergence for COII. Third-position transitions of mt protein-coding genes were saturated (mean Δ value = -0.83 ; table 2).

Discussion

Our results suggest that mt genes have less resolving power than certain nc exons in recovering deep-level mammalian clades. This result was obtained for comparisons between individual mt and nc genes, as well as for comparisons between concatenated nc genes and concatenated mt protein-coding genes from mt genomes. Given that the nc genes exhibited the highest absolute Δ values and that Δ values exceeded intraplacental divergences, it is not surprising that they performed better than mt genes in recovering deep-level clades. We would not expect all types of nc sequences to perform as well as the sequences that were included in our study. Clearly, the exonic sequences included in our study constitute a biased sample from the nc genome. Nevertheless, there is almost unlimited potential for choosing exonic regions with appropriate rates of sequence evolution that allow for the recovery of deep-level mammalian clades. Even though the nc exons performed better than mt genes in recovering deep-level clades, phylogenetic analyses based on mt sequences remain valuable. Complete mt genomes contain considerably more resolving power than single mt genes and have provided strong support for some of the same deep-level clades that are supported by nc sequences, e.g., Hippopotamidae + Cetacea (Ursing and Arnason 1998; Gatesy et al. 1999). Indeed, our comparisons were based on the variable-length bootstrap, which was employed for purposes of evaluating mt and nc sequences on a per-residue basis. Even in our comparisons between concatenated sequences from the mt and nc genomes, the variable-length bootstrap allowed for resampling only at a maximum size of 5,424 bp (fig. 3). Given that mt protein-coding sequences from complete genomes are approximately twice this length and that mt protein-coding sequences are sometimes analyzed in conjunction with mt rRNA sequences, mt genomes provide more resolving power than is evident from figs. 2 and 3 (e.g., see Reyes et al. 2000). The value of mt sequences in phylogenetic analyses is further enhanced if they are collected in conjunction with nc sequences, because they provide an independent estimate of phylogenetic relationships that can be compared with estimates based on nc sequences.

The hypothesis of Arnason, Gullberg, and Janke (1999) that nc alignments must be longer than mt alignments to achieve equivalent levels of resolution is strongly contradicted for the nc exons that we examined. Rather, nc exons recovered benchmark clades with high-

er efficiency. This result suggests that internodal divergence times for the clades considered in this study are not excessively short compared with coalescence times for nc genes. Nuclear genes included in our study represent several unlinked loci, yet each gene individually supports the same widely accepted clades. The conclusion that these nc genes are superior to mt genes in recovering expected relationships does not appear to be the result of selecting poorly performing mt genes. Zardoya and Meyer (1996) classified mt protein-coding genes into three groups (good, medium, and poor) based on their performance in recovering phylogenetically distant relatives. The mt protein-coding genes incorporated in our analysis included one gene (cytochrome *b*) in the good group and two genes (COII, ND1) in the medium group. Furthermore, our comparisons between concatenated mt and nc genes included 12 of 13 protein-coding genes from the mt genome (ND6 was excluded because it is coded on the light strand and has properties different from those of the other 12 protein-coding genes; Waddell et al. 1999).

One factor that may contribute to the differences in performance among the sequences we examined is the rate of nucleotide substitution. Based on the intraplacental divergences reported in table 2, overall rates of nucleotide substitution are highest among the mt protein-coding genes, followed by vWF, IRBP, rRNA, and A2AB. A consequence of the higher substitution rates among the mt protein-coding genes is saturation over shorter lookback times. Less among-sites rate variation among the nc genes may also contribute to differences in performance between the nc and the mt genes. For the 32 pairwise comparisons we performed, maximum-likelihood estimates of the shape parameter (α) for the gamma distribution were estimated with and without an allowance for invariant sites (I). Without an allowance for I, α was higher for the nc gene than for the mt gene in every comparison. With an allowance for I, α was higher for the nc gene in 31 of 32 comparisons (α for mt RNA was higher than α for aquaporin in the mt RNA–aquaporin comparison). With or without I, it appears that substitutions are more evenly distributed among the nc genes and there is less of a tendency for substitutions to occur repeatedly at the same sites.

Mitochondrial protein-coding genes have until now attracted most attention in studies of higher-level relationships for some groups, e.g., classes of vertebrates. Our results suggest that select nc genes may be more appropriate for reconstructing phylogenetic relationships among higher-level taxa. The use of nc sequences to examine phylogenetic relationships is especially important in cases where mt protein-coding genes have suggested hypotheses that do not accord with traditional views (Curolle and Kocher 1999).

Acknowledgments

This work was supported by grants from the National Science Foundation (DEB-9419617 and DEB-9903810 to M.S.S.) and the Nuffield Foundation and the Royal Society (to M.J.S.) and a European Commission–

TMR grant (ERB-FMRX-CT98-0221) to M.J.S. and W.W.d.J. We thank two anonymous referees for helpful comments on an earlier version of this manuscript. Dr. Ulfur Arnason kindly supplied mt alignments.

LITERATURE CITED

- ARNASON, U., and A. GULLBERG. 1996. Cytochrome *b* nucleotide sequences and the identification of five primary lineages of extant cetaceans. *Mol. Biol. Evol.* **13**:407–417.
- 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.
- . 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.
- ASHER, R. J. 1999. A morphological basis for assessing the phylogeny of the “Tenrecoidea” (Mammalia, Lipotyphla). *Cladistics* **15**:231–252.
- CAO, Y., J. ADACHI, A. JANKE, S. PÄÄBO, and M. HASEGAWA. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *J. Mol. Evol.* **39**:519–527.
- CUMMINGS, M. P., S. P. OTTO, and J. WAKELEY. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* **12**:814–822.
- CUROLE, J. P., and T. D. KOCHER. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol. Evol.* **14**:394–398.
- DE JONG, W. 1998. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* **13**:270–275.
- D’ERCHIA, A. M., C. GISSI, G. PESOLE, C. SACCONI, and U. ARNASON. 1996. The guinea pig is not a rodent. *Nature* **381**:597–600.
- GARLAND, T., A. W. DICKERMAN JR., C. M. JANIS, and J. A. JONES. 1993. Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* **42**:265–292.
- GATESY, J., M. MILINKOVITICH, V. WADDELL, and M. J. STANHOPE. 1999. Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. *Syst. Biol.* **48**:6–20.
- GRAUR, D., W. A. HIDE, and H.-W. LI. 1991. Is the guinea-pig a rodent? *Nature* **351**:649–652.
- GRAUR, D., and D. G. HIGGINS. 1994. Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. *Mol. Biol. Evol.* **11**:357–364.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21**:160–174.
- JANKE, A., A. XU, and U. ARNASON. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia and Eutheria. *Proc. Natl. Acad. Sci. USA* **94**:1276–1281.
- KIRSCH, J. A. W., T. F. FLANNERY, M. S. SPRINGER, and F.-J. LAPOINTE. 1995. Phylogeny of the Pteropodidae (Mammalia: Chiroptera) based on DNA hybridisation, with evidence for bat monophyly. *Aust. J. Zool.* **43**:395–428.
- KIRSCH, J. A. W., F.-J. LAPOINTE, and M. S. SPRINGER. 1997. DNA-hybridisation studies of marsupials and their implications for metatherian classification. *Aust. J. Zool.* **45**:211–280.
- 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 the Lipotyphla. *J. Mol. Evol.* **41**:952–957.
- KUMAR, S., and S. B. HEDGES. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**:917–920.
- LAVERGNE, A., E. DOUZERY, T. STICHLER, F. M. CATZEFLIS, and M. S. SPRINGER. 1996. Interordinal mammalian relationships: evidence for paenungulate monophyly is provided by complete mitochondrial 12S rRNA sequences. *Mol. Phylogenet. Evol.* **6**:245–258.
- LIU, R.-G., and M. M. MIYAMOTO. 1999. Phylogenetic assessment of molecular and morphological data for eutherian mammals. *Syst. Biol.* **48**:54–64.
- LOCKHART, P. J., M. A. STEEL, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**:605–612.
- McKENNA, M. C., and S. K. BELL. 1997. Classification of mammals above the species level. Columbia University Press, New York.
- MADSEN, O., P. M. T. DEEN, G. PESOLE, C. SACCONI, and W. W. DE JONG. 1997. Molecular evolution of mammalian aquaporin-2: further evidence that elephant shrew and aardvark join the Paenungulate clade. *Mol. Biol. Evol.* **14**:363–371.
- MOORE, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**:718–726.
- 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.
- NAYLOR, G. J. P., and W. M. BROWN. 1997. Structural biology and phylogenetic estimation. *Nature* **388**:527–528.
- NOVACEK, M. J. 1992. Mammalian phylogeny: shaking the tree. *Nature* **356**:121–125.
- NOWAK, R. M., and J. L. PARADISO. 1983. Walker’s mammals of the world. Johns Hopkins University Press, Baltimore.
- O’LEARY, M. A., and J. H. GEISLER. 1999. The position of Cetacea within Mammalia: phylogenetic analysis of morphological data from extinct and extant taxa. *Syst. Biol.* **48**:455–490.
- PENNY, D., M. HASEGAWA, P. J. WADDELL, and M. D. HENDY. 1999. Mammalian evolution: timing and implications from using the LogDeterminant transform for proteins of differing amino acid composition. *Syst. Biol.* **48**:76–93.
- PETTIGREW, J. D. 1986. Flying primates? Megabats have the advanced pathway from eye to midbrain. *Science* **231**:1304–1306.
- PORTER, C. A., M. GOODMAN, and M. J. STANHOPE. 1996. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. *Mol. Phylogenet. Evol.* **5**:89–101.
- PROTHERO, D. R., and R. M. SCHOCH. 1989. Origin and evolution of the Perissodactyla: summary and synthesis. Pp. 504–529 in D. R. PROTHERO and R. M. SCHOCH, eds. The evolution of perissodactyls. Oxford University Press, Oxford, England.
- RASMUSSEN, A., A. JANKE, and U. ARNASON. 1998. The mitochondrial DNA molecule of the hagfish (*Myxine glutinosa*) and vertebrate phylogeny. *J. Mol. Evol.* **46**:382–388.
- REYES, A., C. GISSI, G. PESOLE, F. M. CATZEFLIS, and C. SACCONI. 2000. Where do rodents fit? Evidence from the complete mitochondrial genome of *Sciurus vulgaris*. *Mol. Biol. Evol.* **17**:979–983.
- REYES, A., G. PESOLE, and C. SACCONI. 1998. Complete mitochondrial DNA sequence of the fat dormouse, *Glis glis*: further evidence of rodent paraphyly. *Mol. Biol. Evol.* **15**:499–505.

- ROWE, T. 1993. Phylogenetic systematics and the early history of mammals. Pp. 129–145 in F. S. SZALAY, M. J. NOVACEK, and M. C. MCKENNA, eds. *Mammal phylogeny: Mesozoic differentiation, multituberculates, monotremes, early therians, and marsupials*. Vol. 1. Springer, New York.
- SHOSHANI, J., and M. C. MCKENNA. 1998. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Mol. Phylogenet. Evol.* **9**:572–584.
- SIMMONS, N. B. 1994. The case for chiropteran monophyly. *Am. Mus. Novit.* **3103**:1–54.
- SPRINGER, M. S. 1997. Molecular clocks and the timing of the placental and marsupial radiations in relation to the Cretaceous-Tertiary boundary. *J. Mammal. Evol.* **4**:285–302.
- SPRINGER, M. S., H. M. AMRINE, A. BURK, and M. J. STANHOPE. 1999. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition. *Syst. Biol.* **48**:65–75.
- SPRINGER, M. S., A. BURK, J. R. KAVANAGH, V. G. WADDELL, and M. J. STANHOPE. 1997a. The interphotoreceptor retinoid binding protein gene in therian mammals: implications for higher level relationships and evidence for loss of function in the marsupial mole. *Proc. Natl. Acad. Sci. USA* **94**:13754–13759.
- SPRINGER, M. S., G. C. CLEVEN, O. MADSEN, W. W. DE JONG, V. G. WADDELL, H. M. AMRINE, and M. J. STANHOPE. 1997b. Endemic African mammals shake the phylogenetic tree. *Nature* **388**:61–64.
- SPRINGER, M. S., L. J. HOLLAR, and A. BURK. 1995. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. *Mol. Biol. Evol.* **12**:1138–1150.
- SPRINGER, M. S., and J. A. W. KIRSCH. 1991. DNA hybridization, the compression effect, and the radiation of diprotodontian marsupials. *Syst. Zool.* **40**:131–151.
- 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.
- STANHOPE, M. J., O. MADSEN, V. G. WADDELL, G. C. CLEVEN, W. W. DE JONG, and M. S. SPRINGER. 1998a. Highly congruent molecular support for a diverse superordinal clade of endemic African mammals. *Mol. Phylogenet. Evol.* **9**:501–508.
- STANHOPE, M. J., M. R. SMITH, V. G. WADDELL, C. A. PORTER, M. S. SHIVJI, and M. GOODMAN. 1996. Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *J. Mol. Evol.* **43**:83–92.
- STANHOPE, M. J., V. G. WADDELL, O. MADSEN, W. W. DE JONG, S. B. HEDGES, G. C. CLEVEN, D. KAO, and M. S. SPRINGER. 1998b. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl. Acad. Sci. USA* **95**:9967–9972.
- SWOFFORD, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.
- 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., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- URSING, B. M., and U. ARNASON. 1998. Analysis of mitochondrial genomes strongly support a hippopotamus-whale clade. *Proc. R. Soc. Lond. B Biol. Sci.* **265**:2251–2255.
- VAWTER, L., and W. M. BROWN. 1986. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* **234**:194–196.
- 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.
- WADDELL, P. J., N. OKADA, and M. HASEGAWA. 1999. Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48**:1–5.
- WIBLE, J. R., and M. J. NOVACEK. 1988. Cranial evidence for the monophyletic origin of bats. *Am. Mus. Novit.* **2911**:1–19.
- WOODBURNE, M. O., and J. A. CASE. 1996. Dispersal, vicariance, and the Late Cretaceous to Early Tertiary land mammal biogeography from South America to Australia. *J. Mammal. Evol.* **3**:121–161.
- WOODBURNE, M. O., B. J. MACFADDEN, J. A. CASE, M. SPRINGER, N. S. PLEDGE, J. D. POWER, J. M. WOODBURNE, and K. JOHNSON. 1993. Land mammal biostratigraphy and magnetostratigraphy of the Etadunna Formation (late Oligocene) of South Australia. *J. Vertebr. Paleontol.* **13**:132–164.
- ZARDOYA, R., and A. MEYER. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Biol. Evol.* **13**:933–942.
- . 1997. The complete DNA sequence of the mitochondrial genome of a “living fossil”, the coelacanth (*Latimeria chalumnae*). *Genetics* **146**:995–1010.

APPENDIX

Taxonomic Composition of Pairwise Comparisons and Benchmark Clades (BC)

- A2AB vs. RNA (31 taxa; data sets resampled at 1,164 bp):** Amblysomus; Balaenoptera; Bos; Bradypus; Cavia; Cynocephalus; Cynopterus/Pteropus; Dicerops/Ceratotherium; Didelphis; Dugong; Echinops; Elephas/Loxodonta; Equus; Erinaceus; Felis; Hippopotamus; Homo; Macropus; Macroscelides/Elephantulus; Macrotus/Tonatia; Manis; Mus; Nycticebus/Galago; Orycteropus; Oryctolagus; Phoca; Procavia; Rattus; Sus; Talpa/Scalopus; Tupaia. BC = CA, CE, CH, EU, PA, PE, and PR.
- A2AB vs. COII (27 taxa; data sets resampled at 675 bp):** Amblysomus; Balaenoptera; Bos; Bradypus/Dasybus; Cavia; Cynocephalus; Cynopterus/Rousettus; Dicerops/Ceratotherium; Didelphis; Elephas/Loxodonta; Equus; Erinaceus; Felis; Hippopotamus; Homo; Macropus; Macroscelides/Elephantulus; Macrotus/Artibeus; Mus; Nycticebus; Orycteropus; Oryctolagus; Phoca; Rattus; Sus; Talpa; Tupaia. BC = CA, CE, CH, EU, PE, and PR.
- A2AB vs. Cytochrome b (24 taxa; data sets resampled at 1,152 bp):** Balaenoptera; Bos; Bradypus/Dasybus; Cavia; Dicerops/Ceratotherium; Didelphis; Dugong; Elephas; Equus; Erinaceus; Felis;

- Hippopotamus; Homo; Macropus; Macroscelides/ Elephantulus; Macrotus/Artibeus; Mus; Orycteropus; Oryctolagus; Phoca; Procavia; Rattus; Sus; Talpa. BC = CA, CE, EU, PA, and PE.
4. **A2AB vs. ND1 (23 taxa; data sets resampled at 957 bp):** Balaenoptera; Bos; Bradypus; Cavia; Dicerops/Ceratotherium; Didelphis; Elephas/Loxodonta; Equus; Erinaceus; Felis; Hippopotamus; Homo; Macropus; Macrotus/Artibeus; Manis; Mus; Nycticebus/Lemur; Orycteropus; Oryctolagus; Phoca; Rattus; Sus; Talpa. BC = CA, CE, EU, PE, and PR.
 5. **IRBP vs. RNA (35 taxa; data sets resampled at 1,292 bp):** Balaenoptera; Bos; Bradypus; Caenolestes; Cynocephalus; Didelphis; Dromiciops; Dugong; Echymipera; Elephantulus; Equus; Erinaceus; Erithizon/Hydrochaeris; Felis; Galago; Hippopotamus; Homo; Loxodonta; Manis; Mus; Notoryctes; Orycteropus; Oryctolagus; Phascogale; Procavia; Pseudocheirus; Pteropus; Sorex; Sus; Tapirus/Ceratotherium; Tonatia; Trichechus; Tupaia; Vulpes/Canis; Vombatus. BC = CA, CE, CH, DI, EU, PA, PE, and PR.
 6. **IRBP vs. COII (26 taxa; data sets resampled at 675 bp):** Balaenoptera; Bos; Bradypus/Dasyurus; Vulpes/Canis; Cynocephalus; Didelphis; Elephantulus; Equus; Erinaceus; Erithizon/Cavia; Felis; Galago; Hippopotamus; Homo; Loxodonta; Mus; Orycteropus; Oryctolagus; Pteropus/Rousettus; Sus; Tapirus/Ceratotherium; Tarsius; Talpa; Tonatia/Artibeus; Tupaia; Vombatus/Macropus. BC = CA, CE, CH, EU, PE, and PR.
 7. **IRBP vs. Cytochrome *b* (32 taxa; data sets resampled at 1,152 bp):** Balaenoptera; Bos; Bradypus/Dasyurus; Caenolestes; Vulpes/Canis; Didelphis; Dromiciops; Dugong; Echymipera; Elephantulus; Equus; Erinaceus; Erithizon/Cavia; Felis; Hippopotamus; Homo; Lama; Loxodonta; Mus; Notoryctes; Orycteropus; Oryctolagus; Phascogale; Procavia; Pseudocheirus; Sorex; Steno/Stenella; Sus; Tapirus/Ceratotherium; Tonatia/Artibeus; Trichechus; Vombatus. BC = CA, CE, CET, DI, EU, PA, and PE.
 8. **IRBP vs. ND1 (25 taxa; data sets resampled at 957 bp):** Balaenoptera; Bos; Bradypus; Vulpes/Canis; Didelphis; Echymipera/Perameles; Equus; Erinaceus; Erithizon/Cavia; Felis; Galago/Lemur; Hippopotamus; Homo; Loxodonta; Manis; Mus; Notoryctes; Orycteropus; Oryctolagus; Phascogale/Sminthopsis; Sus; Talpa; Tapirus/Ceratotherium; Tonatia/Artibeus; Vombatus. BC = CA, CE, EU, PE, and PR.
 9. **vWF vs. RNA (32 taxa; data sets resampled at 1,251 bp):** Amblysomus; Bos; Bradypus; Canis; Ceratotherium; Chaetophractus; Cynocephalus; Dasyprocta/Hydrochaeris; Didelphis; Dobsonia; Dugong; Echinops; Elephantulus; Equus; Erinaceus; Felis; Galago; Hippopotamus; Homo; Loxodonta; Macropus; Manis; Megaderma; Megaptera/Balaenoptera; Orycteropus; Oryctolagus; Procavia; Rattus; Scalopus; Sus; Tonatia; Tupaia. BC = CA, CE, CH, EU, PA, PE, PR, and XE.
 10. **vWF vs. COII (26 taxa; data sets resampled at 675 bp):** Amblysomus; Bos; Canis; Ceratotherium; Chaetophractus/Dasyurus; Cynocephalus; Dasyprocta/Cavia; Didelphis; Dobsonia/Rousettus; Elephantulus; Equus; Erinaceus; Felis; Galago; Hippopotamus; Homo; Loxodonta; Macropus; Megaptera/Balaenoptera; Orycteropus; Oryctolagus; Rattus; Sus; Talpa; Tonatia/Artibeus; Tupaia. BC = CA, CE, CH, EU, PE, and PR.
 11. **vWF vs. Cytochrome *b* (26 taxa; data sets resampled at 1,152 bp):** Bos; Canis; Ceratotherium; Chaetophractus/Dasyurus; Dasyprocta/Cavia; Didelphis; Dugong; Elephantulus; Elephas; Equus; Erinaceus; Felis; Hippopotamus; Homo; Lama; Loxodonta; Macropus; Megaptera/Balaenoptera; Oryctolagus; Orycteropus; Phocoena; Procavia; Rattus; Scalopus/Talpa; Sus; Tonatia/Artibeus. BC = CA, CE, CET, EU, PA, and PE.
 12. **vWF vs. ND1 (23 taxa; data sets resampled at 957 bp):** Bos; Bradypus; Canis; Ceratotherium; Chaetophractus/Dasyurus; Dasyprocta/Cavia; Didelphis; Equus; Erinaceus; Felis; Galago/Lemur; Hippopotamus; Homo; Loxodonta; Macropus; Manis; Megaptera/Balaenoptera; Orycteropus; Oryctolagus; Rattus; Scalopus/Talpa; Sus; Tonatia/Artibeus. BC = CA, CE, EU, PE, PR, and XE.
 13. **γ -fibrinogen vs. RNA (23 taxa; data sets resampled at 679 bp):** Antilocapra; Balaenoptera; Bos; Camelus; Canis; Dasyurus; Dicerorhinus/Ceratotherium; Dugong; Elephantulus; Equus; Felis; Hexaprotodon/Hippopotamus; Homo; Loxodonta; Manis; Mus; Orycteropus; Ovis; Procavia; Rattus; Sus; Tragulus; Trichechus. BC = CA, CE, PA, PE, and RU.
 14. **γ -fibrinogen vs. COII (17 taxa; data sets resampled at 675 bp):** Balaenoptera; Bos; Canis; Dasyurus; Dicerorhinus/Ceratotherium; Elephantulus; Equus; Felis; Hexaprotodon/Hippopotamus; Homo; Loxodonta; Mus; Orycteropus; Ovis; Rattus; Saimiri; Sus. BC = CA, CE, and PE.
 15. **γ -fibrinogen vs. Cytochrome *b* (29 taxa; resampled at 679 bp):** Antilocapra; Babyrousa; Balaenoptera; Bos; Camelus; Canis; Dasyurus; Delphinapterus; Dicerops; Dugong; Elephantulus; Equus; Felis; Giraffa; Hexaprotodon; Homo; Loxodonta; Mus; Orycteropus; Ovis; Physeter; Procavia; Rattus; Saimiri; Sus; Tayassu; Tragulus; Trichechus; Ziphilus. BC = CA, CE, CET, PA, PE, RU, and SU.
 16. **γ -fibrinogen vs. ND1 (15 taxa; resampled at 679 bp):** Balaenoptera; Bos; Canis; Dasyurus; Dicerops; Equus; Felis; Homo; Loxodonta; Manis; Mus; Orycteropus; Ovis; Rattus; Sus. BC = CA, CE, and PE.
 17. **β -casein vs. RNA (27 taxa; resampled at 520 bp):** Aepyceros; Antilocapra; Balaenoptera; Bos; Bose-laphus; Camelus; Capra; Cervus; Damaliscus; Equus; Gazella; Hippopotamus; Homo; Kobus; Muntiacus; Mus; Orycteropus; Oryctolagus; Oryx; Ovis; Panthera/Felis; Phoca; Rattus; Rhinoceros; Sus; Tragelaphus; Tragulus. BC = CA, CE, PE, and RU.

18. **β -casein vs. COII (14 taxa; resampled at 520 bp):** Balaenoptera; Bos; Equus; Hippopotamus; Homo; Mus; Orycteropus; Oryctolagus; Ovis; Panthera/Felis; Phoca; Rattus; Rhinoceros; Sus. BC = CA, CE, and PE.
19. **β -casein vs. Cytochrome *b* (35 taxa; resampled at 520 bp):** Ailurus; Antilocapra; Babyrousa; Balaenoptera; Bos; Bubalus; Camelus; Capra; Cervus; Delphinapterus; Dicerus; Equus; Giraffa; Hexaprotodon; Hippopotamus; Homo; Lagenorhynchus; Megaptera; Mesoeux; Monodon; Mus; Orycteropus; Oryctolagus; Ovibos; Ovis; Panthera/Felis; Phoca; Physeter; Procyon; Rattus; Sus; Syncerus; Tayassu; Tragulus; Ziphilus. BC = CA, CE, CET, PE, RU, and SU.
20. **β -casein vs. ND1 (14 taxa; resampled at 520 bp):** Balaenoptera; Bos; Equus; Hippopotamus; Homo; Mus; Orycteropus; Oryctolagus; Ovis; Panthera/Felis; Phoca; Rattus; Rhinoceros; Sus. BC = CA, CE, and PE.
21. **κ -casein vs. RNA (26 taxa; resampled at 519 bp):** Antilocapra; Balaenoptera; Bos; Boselaphus; Camelus; Canis; Capra; Cavia; Cervus; Equus; Hippopotamus; Homo; Kobus; Lama; Muntiacus; Mus; Orycteropus; Oryctolagus; Oryx; Ovis; Panthera/Felis; Rattus; Rhinoceros; Sus; Tragelaphus; Tragulus. BC = CA, CE, PE, and RU.
22. **κ -casein vs. COII (15 taxa; resampled at 519 bp):** Balaenoptera; Bos; Canis; Cavia; Dicerorhinus/Ceratotherium; Equus; Hippopotamus; Homo; Mus; Orycteropus; Oryctolagus; Ovis; Panthera/Felis; Rattus; Sus. BC = CA, CE, and PE.
23. **κ -casein vs. Cytochrome *b* (36 taxa; resampled at 519 bp):** Ailurus; Antilocapra; Babyrousa; Balaenoptera; Bos; Bubalus (*B. bubalis* and *B. depressicornis*); Camelus (*C. bactrianus* and *C. dromedarius*); Canis; Capra; Capracornis; Cavia; Cervus; Equus; Giraffa; Hexaprotodon; Hippopotamus; Homo; Lagenorhynchus; Lama; Mus; Nemorhaedus; Orycteropus; Oryctolagus; Ovis; Panthera/Felis; Physeter; Procyon; Rattus; Rhinoceros; Sus; Syncerus; Tayassu; Tragulus; Ziphilus. BC = CA, CE, CET, PE, RU, and SU.
24. **κ -casein vs. ND1 (15 taxa; resampled at 519 bp):** Balaenoptera; Bos; Canis; Cavia; Equus; Hippopotamus; Homo; Mus; Orycteropus; Oryctolagus; Ovis; Panthera/Felis; Rattus; Rhinoceros; Sus. BC = CA, CE, and PE.
25. **Protamine vs. RNA (18 taxa; resampled at 386 bp):** Balaenoptera; Bos; Camelus; Cavia; Cervus; Dicerus/Rhinoceros; Elephas/Loxodonta; Equus; Felis; Gazella; Hexaprotodon/Hippopotamus; Homo; Lama; Mus; Rattus; Sus; Tragulus; Ursus/Canis. BC = CA, CE, PE, and RU.
26. **Protamine vs. COII (14 taxa; resampled at 386 bp):** Balaenoptera; Bos; Cavia; Dicerus/Rhinoceros; Elephas/Loxodonta; Equus; Felis; Hexaprotodon/Hippopotamus; Homo; Mus; Rattus; Saimiri; Sus; Ursus/Canis. BC = CA, CE, and PE.
27. **Protamine vs. Cytochrome *b* (23 taxa; resampled at 386 bp):** Balaenoptera; Bos; Camelus; Cavia; Cervus; Delphinapterus; Dicerus/Rhinoceros; Elephas/Loxodonta; Equus; Felis; Giraffa; Hexaprotodon/Hippopotamus; Homo; Lagenorhynchus; Lama; Mus; Orcinus; Rattus; Saimiri; Sus; Tayassu; Tragulus; Ursus/Canis. BC = CA, CE, CET, PE, RU, and SU.
28. **Protamine vs. ND1 (13 taxa; resampled at 386 bp):** Balaenoptera; Bos; Cavia; Dicerus/Rhinoceros; Elephas/Loxodonta; Equus; Felis; Hexaprotodon/Hippopotamus; Homo; Mus; Rattus; Sus; Ursus/Canis. BC = CA, CE, and PE.
29. **Aquaporin vs. RNA (23 taxa; resampled at 333 bp):** Amblysomus; Balaenoptera; Bos; Cavia; Canis; Dasypus; Dugong; Echinops; Elephas/Loxodonta; Equus; Erinaceus; Homo; Macroscelides/Elephantulus; Manis; Mus; Orycteropus; Oryctolagus; Ovis; Procavia; Rattus; Sus; Talpa; Trichechus. BC = CE and PA.
30. **Aquaporin vs. COII (18 taxa; resampled at 333 bp):** Amblysomus; Balaenoptera; Bos; Canis; Cavia; Dasypus; Elephas/Loxodonta; Equus; Erinaceus; Homo; Macroscelides/Elephantulus; Mus; Orycteropus; Oryctolagus; Ovis; Rattus; Sus; Talpa. BC = CE.
31. **Aquaporin vs. Cytochrome *b* (21 taxa; resampled at 333 bp):** Balaenoptera; Bos; Canis; Cavia; Dasypus; Dugong; Elephas/Loxodonta; Equus; Erinaceus; Homo; Macroscelides/Elephantulus; Mus; Orycteropus; Oryctolagus; Ovis; Physeter; Procavia; Rattus; Sus; Talpa; Trichechus. BC = CE, CET, and PA.
32. **Aquaporin vs. ND1 (17 taxa; resampled at 333 bp):** Balaenoptera; Bos; Canis; Cavia; Dasypus; Elephas/Loxodonta; Equus; Erinaceus; Homo; Manis; Mus; Orycteropus; Oryctolagus; Ovis; Rattus; Sus; Talpa. BC = CE.

NOTE.—Slashes indicate chimerics; the taxon before the slash is the nuclear sequence, and the taxon after the slash is the mitochondrial sequence. Abbreviations for benchmark comparisons: CA = Carnivora; CE = Cetartiodactyla; CET = Cetacea; CH = Chiroptera; DI = Diprotodontia; EU = Eutheria; PA = Paenungulata; PE = Perissodactyla; PR = Primates; RU = Ruminantia; SU = Suina; XE = Xenarthra.

DAN GRAUR, reviewing editor

Accepted September 26, 2000