ACCEPTED FOR PUBLICATION IN
Molecular Phylogenetics and Evolution (2001) vol. 21, issue # 1, pp. 43–54.

THIS DOCUMENT IS IDENTICAL TO CORRECTED PROOF (FINAL PUBLISHED VERSION)

http://www.sciencedirect.com/science/article/pii/S1055790301909893
http://pubmed.gov/11603936

Sources of Incongruence among Mammalian Mitochondrial Sequences: COII, COIII, and ND6 Genes Are Main Contributors

Nikolai A. Shevchuk* and
Marc W. Allard†

**Institute for Biomedical Sciences and †Department of Biological Sciences, The George Washington University, Washington, DC 20052*

Received September 15, 2000;
revised March 30, 2001

To investigate the origins of incongruence among mammalian mitochondrial protein-coding genes, we compiled a matrix that included 13 protein-coding genes for 41 mammals from 14 different orders. This matrix was examined for congruence using different partitioning strategies. The incongruence length difference test showed significant incongruence among the 13 gene partitions used simultaneously, and the result was not affected by third codon or transversion weighting. In the pair-wise comparisons, significant incongruence was detected between NADH:ubiquinone oxidoreductase subunit 6 gene (ND6), cytochrome oxidase subunit II (COII), or cytochrome oxidase subunit III (COIII) gene partitioned individually against the rest of the genes. Omission of any of the 14 mammalian orders alone or in combinations from the matrix did not result in a statistically significant improvement of congruence, suggesting that taxonomic sampling will not

improve congruence among the data sets. However, omission of the ND6, COII, and COIII significantly improved congruence in our data matrix. Possible origins of unusual phylogenetic properties of the three genes are discussed. ©2001 Academic Press

Key Words: **mitochondrial DNA; mammalian phylogeny; incongruence; weighting; protein-coding genes.**

INTRODUCTION

A number of studies in evolutionary biology use phylogenetic information obtained from mitochondrial DNA. Researchers expect mtDNA to be a generally reliable marker for addressing various questions in population genetics and systematics. However, regardless of the phylogenetic method and weighting schemes used, some mitochondrial genes can lead to highly unconventional evolutionary conclusions, which disagree with whole mitochondrial genome data and thus may not accurately reflect history of the organisms in question (Naylor and Brown, 1998). Topological congruence between phylogenies produced by independent data partitions is thought to constitute a strong support for phylogenetic relationships (Miyamoto and Fitch, 1995) and hence the relationship between congruence among trees and data sets and possibly phylogenetic accuracy may be key in a cladistic analysis. In this report, we discuss how one large set of data matrices is affected by various partitioning and weighting strategies, and we believe our results offer new ways of analyzing characters and support our previous conclusions about the weighting and combining of multiple data sets.

There are several available methods of measuring the agreement within and between data sets, e.g., measures of consistency, such as the consistency index (CI) and the retention index (RI) (Farris, 1989), and congruence, such as the

incongruence length difference (ILD) test. The ILD test is currently available in such computer programs as Xarn (Farris *et al.*, 1994), Winclada (Nixon, 2000), and PAUP* (Swofford, 2000). The ILD test (also called partition homogeneity test in PAUP*) compares the number of steps required for minimum-length trees in separate and combined analyses. In this test, between-partition incongruence is measured by the additional steps required when the data are combined in a simultaneous analysis (Farris *et al.*, 1994). The ILD test can be applied to any number of data partitions and can be applied under a wide variety of weighting methods.

Here, we test whether various methods including weighting strategies can improve the congruence between mammalian mitochondrial sequence data by measuring the incongruence length difference using different ways of partitioning, both by genes and by taxonomic sampling of the matrix that consisted of 13 protein-coding genes for 41 mammals from 14 different orders of placental mammals. If various partitioning and sampling strategies are removing heterogeneity from molecular data sets, then we hypothesize that these altered matrices should show improved values of congruence. If no differences are measured between the alternative sampling strategies, this then suggests that there has been no removal of the heterogeneity in these mammalian mitochondrial sequences and that no single gene or taxon is contributing to overall incongruence. By careful dissection of partitions, one may be able to locate where incongruence resides in a large combined matrix (DeSalle and Brower, 1997; Remsen and DeSalle, 1998; Thornton and DeSalle, 2000).

MATERIALS AND METHODS

We constructed data matrices for 13 protein-coding genes for 41 mammals (34 species) (Table 1). The complete mitochondrial genome for each of these taxa was obtained from the GenBank (NCBI) with the MacVector program (Eastman Kodak Co.). We downloaded the specific genes using the features table listed in each GenBank data file. We determined alignments with reference to the amino acid code, which was translated from the original sequence. All matrices were constructed twice independently and each was analyzed with all characters equally weighted and unordered. We used tree length and shape to assess whether there were editorial errors in the final matrices. We assumed that this would reduce editing errors, as independent results should be identical. We then combined all genes into a single matrix which included the 13 protein-coding genes for 41 mammals, 11,463 characters each, for about half a million base pairs of mitochondrial sequence. This matrix was examined for 14 different partitioning strategies including 1 test with 13 partitions corresponding to each gene and 13 tests with each gene partitioned against the rest of the genes. We also examined the matrix for omissions of some groups and separate taxa and for three separate weighting strategies including equal weighting, transversion weighting (i.e., recognition of two states, either purine or pyrimidine), and codon weighting (in which the third codon position was omitted).

TABLE 1

Taxa sampled for complete mtDNA genomes including accession numbers, species, and common name.

Order	Genebank Accession #	Species	Common name
Artiodactyla[a]	AJ010957	*Hippopotamus amphibius*	Hippopotamus
	AF034253	*Sus scrofa*	Pig
	AJ002189	*Sus scrofa 2*	Pig
	AF010406	*Ovis aries*	Sheep
	V00654	*Bos taurus*	Cattle
Perrisodactyla[b]	X79547	*Equus caballus*	Horse
	X97337	*Equus asinus*	Donkey
	X97336	*Rhinoceros unicornis*	Indian rhino
	Y07726	*Ceratotherium simum*	White rhino
Cetacea[c]	X61145	*Balaenoptera physalus*	Fin whale
	X72204	*Balaenoptera musculus*	Blue whale
Primates[d]	D38112	*Homo sapiens*	Man
	V00662	*Homo sapiens 2*	Man
	J01415	*Homo sapiens 3*	Man
	X93334	*Homo sapiens 4*	Man
	D38115	*Pongo pygmaeus*	Orangutan
	X97707	*Pongo pygmaeus abelii*	Orangutan
	D38116	*Pan panicus*	Pygmy chimp
	D38113	*Pan troglodytes*	Common chimp
	X93335	*Pan troglodytes 2*	Common chimp
	D38114	*Gorilla gorilla*	Gorilla
	X93347	*Gorilla gorilla 2*	Gorilla
	Y18001	*Papio hamadryas*	Baboon
	X99256	*Hylobates lar*	Gibbon
Rodentia[e]	AJ222767	*Cavia porcellus*	Guinea pig
	AJ001562	*Glis glis*	Fat dormouse

	V00711	*Mus musculus*	Mouse
	X14848	*Rattus norvegicus*	Rat
Carnivora[f]	U96639	*Canis familiaris*	Dog
	X63726	*Phoca vitulina*	Harbor seal
	X72004	*Halichoerus grypus*	Grey seal
	U20753	*Felis catus*	Cat
Insectivora[g]	X88898	*Erinaceus europeus*	Hedgehog
Tubulidentata[h]	Y18475	*Orycteropus afer*	Aardvark
Chiroptera[i]	AF061340	*Artibeus jamaicensis*	Fruit-eating bat
Lagomorpha[j]	AJ001588	*Oryctolagus cuniculus*	Rabbit
Edentata[k]	Y11832	*Dasypus novemcinctus*	Armadillo
Monotremata[l]	X83427	*Ornithorhynchus anatinus*	Platypus
Marsupialia[m]	Z29573	*Didelphis virginiana*	Opossum
	Y10524	*Macropus robustus*	Wallaroo
Proboscidea[n]	AJ224821	*Loxodonta africana*	African elephant

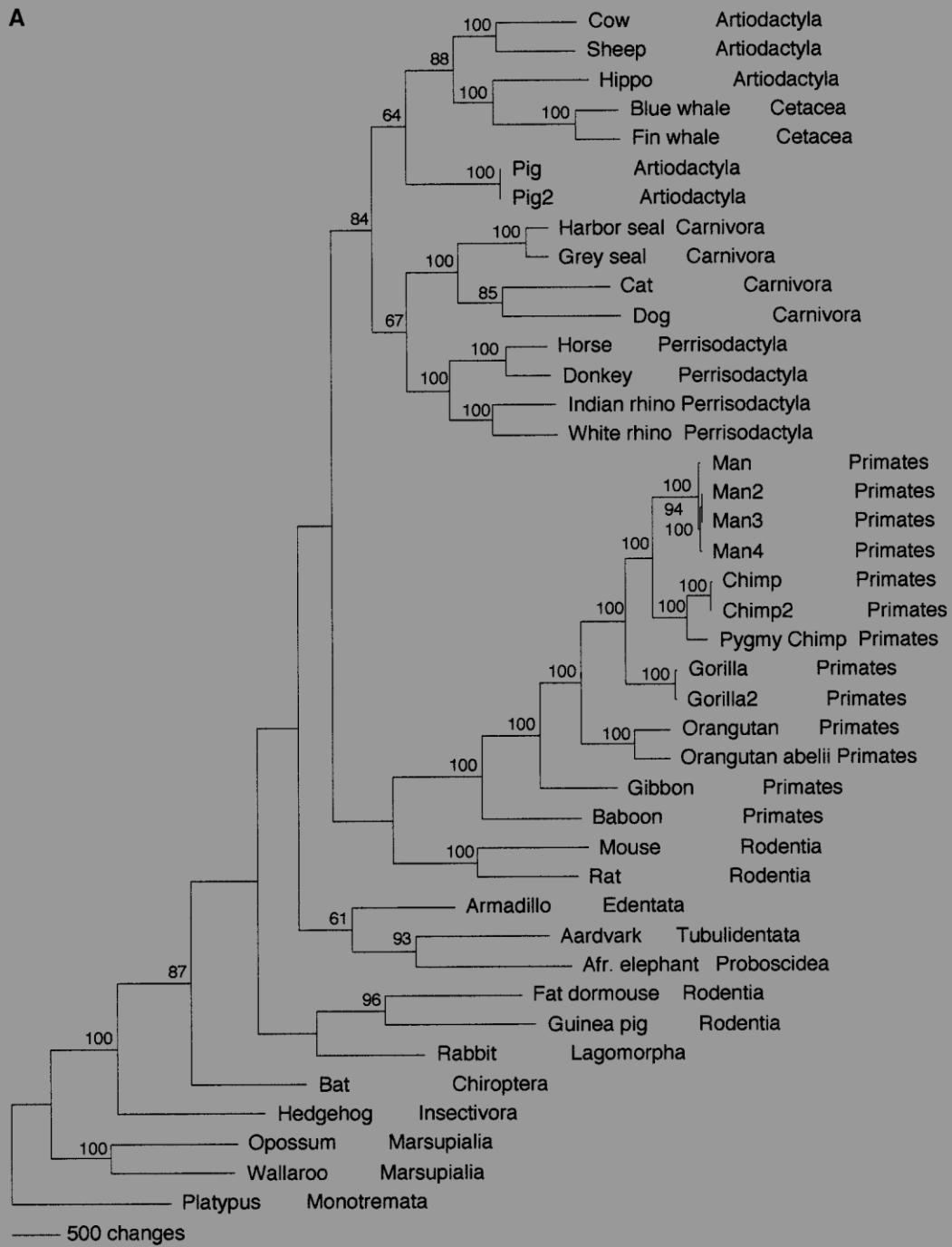
From top to bottom: [a]Ursing and Arnason (1998a), Lin *et al.* (1999), Ursing and Arnason (1998b), Hiendleder *et al.* (1998), Anderson *et al.* (1982), [b]Xu and Arnason (1994), Xu *et al.* (1996a), Xu *et al.* (1996b), Xu and Arnason (1997), [c]Arnason *et al.* (1991), Arnason and Gullberg (1993), [d]Horai *et al.* (1995), Anderson *et al.* (1981), Montoya *et al.* (1981), Arnason *et al.* (1996a), Xu and Arnason (1996a), Xu and Arnason (1996a), Horai *et al.* (1995), Horai *et al.* (1992), Arnason *et al.* (1996a), Xu and Arnason (1996b), Arnason *et al.* (1998), Arnason *et al.* (1996b), [e]D'Erchia *et al.* (1996), Reyes *et al.* (1998), Bibb *et al.* (1981), Gadaleta *et al.* (1989), [f]Kim et al (1998), Arnason and Johnson (1992), Arnason *et al.* (1993), Lopez et al (1996), [g]Krettek *et al.* (1995), [h]Arnason *et al.* (1999), [i]Pumo *et al.* (1998), [j]Gissi *et al.* (1998), [k]Arnason *et al.* (1997), [l]Janke *et al.* (1996), [m]Janke *et al.* (1994), Janke *et al.* (1997), [n]Hauf, J., Waddell, P. J., Chalwatzis, N., Joger, U., Zimmermann, F. K. (1999, unpublished)

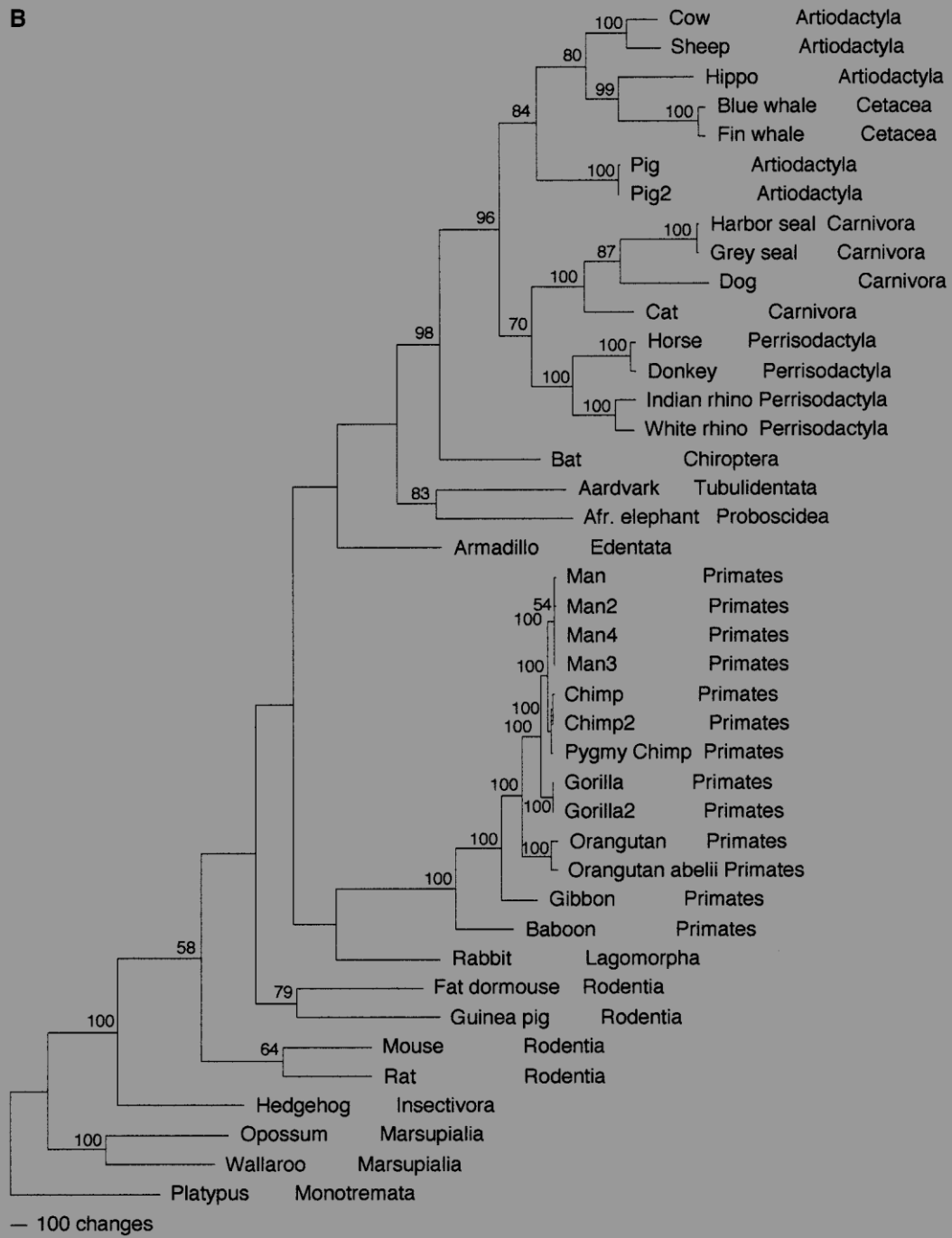
We assessed the genes as to whether they were congruent with one another using the ILD test, which is implemented in the program PAUP* 4.0a2b (Swofford, 2000) and Winclada 0.9.9 (Nixon, 2000). The ILD partitions followed the gene

alignment sequence lengths and included constant and other phylogenetically uninformative positions.

RESULTS AND DISCUSSION

All 41 mammals (34 different taxa) and their respective GenBank accession numbers and abbreviations are listed in Table 1. The resulting combined matrix had the dimensions 41 × 11,463 and included 13 mitochondrial protein-coding genes for each of the 41 mammals. Heuristic search with equal weighting of characters yielded a single most parsimonious tree (Fig. 1A). However, contrary to our expectations (Allard *et al.*, 1999a), transversion and codon weighting did not produce trees identical to the tree corresponding to equal weighting. Analysis with transversion weighting yielded one tree (Fig. 1B), whereas analysis with the third codon omitted gave rise to two shortest trees (Figs. 1C and 1D), which suggests that the third codon positions do contain a phylogenetic signal necessary to get a resolved tree (Kallersjo *et al.*, 1999).


A
Cow Artiodactyla
Sheep Artiodactyla
Hippo Artiodactyla
Blue whale Cetacea
Fin whale Cetacea
Pig Artiodactyla
Pig2 Artiodactyla
Harbor seal Carnivora
Grey seal Carnivora
Cat Carnivora
Dog Carnivora
Horse Perrisodactyla
Donkey Perrisodactyla
Indian rhino Perrisodactyla
White rhino Perrisodactyla
Man Primates
Man2 Primates
Man3 Primates
Man4 Primates
Chimp Primates
Chimp2 Primates
Pygmy Chimp Primates
Gorilla Primates
Gorilla2 Primates
Orangutan Primates
Orangutan abelii Primates
Gibbon Primates
Baboon Primates
Mouse Rodentia
Rat Rodentia
Armadillo Edentata
Aardvark Tubulidentata
Afr. elephant Proboscidea
Fat dormouse Rodentia
Guinea pig Rodentia
Rabbit Lagomorpha
Bat Chiroptera
Hedgehog Insectivora
Opossum Marsupialia
Wallaroo Marsupialia
Platypus Monotremata
500 changes


B
Cow Artiodactyla
Sheep Artiodactyla
Hippo Artiodactyla
Blue whale Cetacea
Fin whale Cetacea
Pig Artiodactyla
Pig2 Artiodactyla
Harbor seal Carnivora
Grey seal Carnivora
Dog Carnivora
Cat Carnivora
Horse Perrisodactyla
Donkey Perrisodactyla
Indian rhino Perrisodactyla
White rhino Perrisodactyla
Bat Chiroptera
Aardvark Tubulidentata
Afr. elephant Proboscidea
Armadillo Edentata
Man Primates
Man2 Primates
Man4 Primates
Man3 Primates
Chimp Primates
Chimp2 Primates
Pygmy Chimp Primates
Gorilla Primates
Gorilla2 Primates
Orangutan Primates
Orangutan abelii Primates
Gibbon Primates
Baboon Primates
Rabbit Lagomorpha
Fat dormouse Rodentia
Guinea pig Rodentia
Mouse Rodentia
Rat Rodentia
Hedgehog Insectivora
Opossum Marsupialia
Wallaroo Marsupialia
Platypus Monotremata
— 100 changes

C

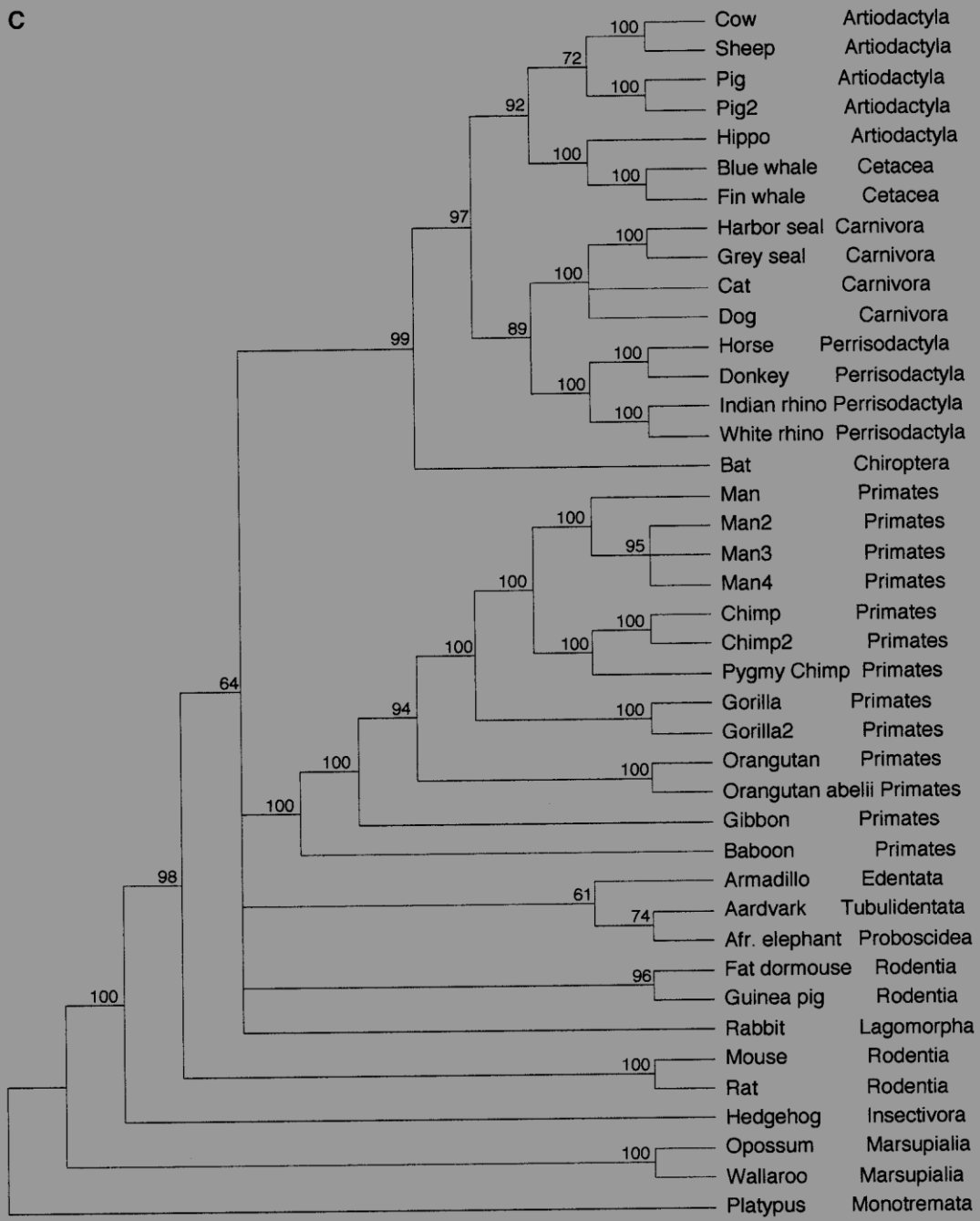

Cow Artiodactyla
Sheep Artiodactyla
Pig Artiodactyla
Pig2 Artiodactyla
Hippo Artiodactyla
Blue whale Cetacea
Fin whale Cetacea
Harbor seal Carnivora
Grey seal Carnivora
Cat Carnivora
Dog Carnivora
Horse Perrisodactyla
Donkey Perrisodactyla
Indian rhino Perrisodactyla
White rhino Perrisodactyla
Bat Chiroptera
Man Primates
Man2 Primates
Man3 Primates
Man4 Primates
Chimp Primates
Chimp2 Primates
Pygmy Chimp Primates
Gorilla Primates
Gorilla2 Primates
Orangutan Primates
Orangutan abelii Primates
Gibbon Primates
Baboon Primates
Armadillo Edentata
Aardvark Tubulidentata
Afr. elephant Proboscidea
Fat dormouse Rodentia
Guinea pig Rodentia
Rabbit Lagomorpha
Mouse Rodentia
Rat Rodentia
Hedgehog Insectivora
Opossum Marsupialia
Wallaroo Marsupialia
Platypus Monotremata

Fig. 1. (A) The most parsimonious tree obtained by simultaneous analysis of 13 mitochondrial genes with equal weighting of characters; (B) the most parsimonious tree obtained by simultaneous analysis of 13 mitochondrial genes with transversion weighting of characters; (C) strict consensus of two most parsimonious trees resulting from simultaneous analysis of 13 mitochondrial genes with third codon position downweighted. Uninformative characters were excluded from analyses. Bootstrap scores are shown near the internal branches. Tree lengths, CI, and RI are listed in table 2. Branch lengths are drawn to scale for figures A and B.

Our most parsimonious tree, which corresponds to equal weighting of combined data (Fig. 1A), is fully consistent with the previously published tree (Allard *et al.*, 1999a). Basic characteristics of all of the trees are listed in Table 2. We rooted the trees in accordance with the common recognition of *Monotremata* (platypus) as one of the most primitive mammals (Novacek, 1992). Neither of the trees completely supported previously published mammalian phylogenies that are based on morphological and fossil data (Novacek, 1992).

TABLE 2

Tree information for the most parsimonious solutions for the three weighting strategies of the combined matrix.

Weighting strategy	Number of trees	Tree length	Total characters	Informative characters	CI	RI
Equal weighting	1	50,258	11,463	6,642	0.27	0.51
Transversions only	1	20,811	11,463	4,327	0.21	0.58
Codon weighting	2	16,823	7,642	3,013	0.32	0.59
Without COII, COIII, and ND6	2	42,027	9,456	5,524	0.26	0.50

Note: The data were obtained using PAUP*, with 1000 random tree bisection-reconnections in the heuristic search option. Uninformative characters were excluded from analyses. A complete alignment is available from **http://www.gwu.edu/~clade/faculty/allard/data.html** and EBI, accession number DS44554

The incongruence length difference test of all the genes analyzed separately by PAUP* led to the conclusion of incongruence (Table 3), which may be caused by sheer numbers of taxa in the new matrix (41), which is about double the size of the matrix analyzed in Allard *et al.* (1999a). An alternative explanation for incongruence could be the influence of hypothetical highly homoplasious genes and/or taxa. To trace the source of incongruence in our data set, we analyzed individual genes partitioned against the rest of the combined matrix using both PAUP* and Winclada. As one can see in Table 3, the results obtained with the two programs are mostly consistent, and several genes show high levels of congruence with the rest of the data. Although there is some variation between the results produced by PAUP* and Winclada, three genes, COII, COIII, and ND6, show incongruence with the rest of the combined data for both programs (Table 3). To evaluate the utility of weighting strategies for improvement of congruence, we applied either transversion or codon weighting to our data set and measured ILD between all the partitions (using only PAUP*). This approach generally improved congruence among the partitions, including the COII and COIII genes (Table 3), but did not have a noticeable effect on all genes analyzed separately and on ND6 partitioned against the rest of the matrix (Table 3).

TABLE 3

Incongruence length differences (ILDs) for 14 distinct partitioning strategies of the combined matrix of mitochondrial protein-coding genes.

Partitioning		PAUP* ILD	Winclada ILD	Transversions only, PAUP* ILD	Codon Weighting, PAUP* ILD
ND1	against the rest	0.20	0.1700	0.92	0.26
ND2	against the rest	0.11	0.5248	0.59	0.90
COI	against the rest	0.08	0.2673	0.02	0.44
COII	against the rest	0.01	0.0495	0.31	0.26
ATP8	against the rest	0.42	0.0730	0.58	0.34
ATP6	against the rest	0.56	0.2178	0.97	0.73
COIII	against the rest	0.03	0.0297	0.23	0.09
ND3	against the rest	0.92	0.8614	0.51	0.88
ND4L	against the rest	0.39	0.3762	0.25	0.55
ND4	against the rest	0.08	0.7030	0.39	0.15
ND5	against the rest	0.49	0.5248	0.18	0.17
ND6	against the rest	0.01	0.0099	0.04	0.01
CytB	against the rest	0.27	0.7624	0.28	0.05
13 genes separately		0.01	-	0.01	0.01

Note: The data were obtained using partition homogeneity (ILD) test implemented in the program PAUP* (settings: 100 heuristic search replicates, 6 branch swappings per each tree bisection, no reconnection limit, saving of multiple trees allowed), and Winclada (settings: 100 replications, 2 mult reps/replication, 2 trees to hold/mult rep, 10 trees to hold, amb-poly=(default)). The current implementation of Winclada can only test two partitions at a time. ILD values less than 0.05 indicate incongruence between given partitions. For the incongruent partitions, we also performed the tests with transversion or codon weighting to evaluate the utility of such approaches for the improvement of congruence.

To further localize incongruence among the mitochondrial protein-coding genes, we performed all possible pairwise analyses for COII, COIII, and ND6 using PAUP* (Table 4). Interestingly, COII and COIII each are incongruent with 5 of the remaining 12 genes, and ND6 is incongruent with all of them. Thus, being incongruent with many other genes, these 3 genes could be a hypothetical source of incongruence among the 13 mammalian protein-coding genes, although, in terms of percentage of informative characters, CI, RI, and number of trees, the topologies derived from these genes analyzed individually were not consistently different from those of the other 10 genes (data not shown).

TABLE 4

Incongruence results for pairwise gene comparisons of the three incongruent genes against the rest

	COII	COIII	ND6
ND1	0.05	0.03	0.01
ND2	0.02	0.02	0.02
COI	0.59	0.89	0.01
COII	-	0.11	0.01
ATP6	0.49	0.03	0.01
ATP8	0.10	0.08	0.01
COIII	0.12	-	0.01
ND3	0.37	0.72	0.01
ND4L	0.02	0.10	0.01
ND4	0.19	0.02	0.01
ND5	0.01	0.16	0.01
ND6	0.01	0.01	-
CytB	0.58	0.27	0.01

Note. Program settings for PAUP* were as follows: 100 replications, TBR, 2 random swappings per replication, 10 trees to hold, no reconnection limit. ILD values less than 0.05 indicate incongruence between gene partitions. Bonferroni correction for type I error problems of multiple testing is 0.05/78 (P value/the number of possible pairwise comparisons) 5 6.41 3 1024. This is compared to the probability that 22 of the pairwise comparisons were obtained by chance alone 1.44 3 10241 (obtained by the multiplication of all significant P values in Table 4). These results are significant even after a Bonferroni correction is applied (M. E. Siddall pers. comm.).

Another potential source of incongruence in sequence data is individual mammalian taxa, whose genes to a greater or lesser extent might contain homoplasious characters or may have evolved under an unusual genetic process (e.g., concerted evolution or horizontal gene transfer). Therefore, deletion of such taxa from the combined matrix could improve congruence among mammalian mitochondrial genes, provided such highly homoplasious species do exist (Lecointre *et al.*, 1998). As one can see in Fig. 2, no one mammalian order or species is apparently responsible for the overall incongruence among mitochondrial protein-coding genes. Although deletion of such clades as the eutherian orders Primates and Carnivora can slightly increase ILD values (Fig. 2), this increase is not statistically significant ($P < 0.05$).

DELETED	PAUP* ILD
NONE	0.01
FIN WHALE	0.01
BLUE WHALE	0.01
COW	0.01
SHEEP	0.01
HIPPOPOTAMUS	0.01
PIG	0.01
PIG2	0.01
HARBOR SEAL	0.03
GREY SEAL	0.02
CAT	0.01
DOG	0.01
HORSE	0.01
DONKEY	0.01
INDIAN RHINO	0.01
WHITE RHINO	0.01
MAN	0.02
MAN2	0.01
MAN3	0.01
MAN4	0.01
CHIMP	0.01
CHIMP2	0.01
PYGMY CHIMP	0.01
GORILLA	0.01
GORILLA2	0.01
ORANGUTAN	0.01
ORANGUTAN abelii	0.01
GIBBON	0.01
BABOON	0.01
ARMADILLO	0.01
AARDVARK	0.01
ELEPHANT	0.01
MOUSE	0.01
RAT	0.01
FAT DORMOUSE	0.01
GUINEA PIG	0.01
RABBIT	0.01
BAT	0.01
HEDGEHOG	0.01
OPOSSUM	0.01
WALLAROO	0.01

Fig. 2. Incongruence length differences (ILDs) for the combined matrix with some of the taxa or clades deleted. Blocks define taxa that were omitted from a test. The data were obtained using partition homogeneity test implemented in the program PAUP*, with 100 heuristic search replicates, 6 branch swappings per each tree bisection, no reconnection limit, saving of multiple trees allowed. ILD values less than 0.05 indicate incongruence between gene partitions. In each test, all 13 genes were partitioned separately (i.e. there were 13 gene partitions).

Since sampling of taxa does not seem to dramatically affect congruence, it was of interest to test whether removal of such genes as ND6, COII, and COIII from the data set could affect the phylogenetic tree and improve congruence. We performed three ILD tests (using PAUP*) corresponding to deletion of each of the three genes independently and one ILD test with all three genes deleted simultaneously (the rest of the genes were analyzed separately in each case, with 12 and 10 partitions, respectively, for the independent and simultaneous deletions). Twelve genes analyzed separately without COII yielded an ILD value of 0.02, 12 genes without COIII yielded a value of 0.01, and 12 genes without ND6 still yielded a value of 0.01. As one can see, no single gene seems to be responsible for all of the incongruence in mitochondrial protein-coding genes, since there is not a significant increase in ILD values following deletion of individual genes compared to the value of 0.01 for all 13 partitions. However, deletion of all 3 genes yields an appreciable improvement of congruence (from 0.01 to 0.18), suggesting that the remaining 10 genes (*ATP8*, *ATP6*, ND1, ND2, ND3, ND4, ND4L, ND5, COI, and CytB) are congruent among the 41 mammals (34 species) analyzed here. Interestingly, the 3 anomalous genes have such a dramatic effect on congruence notwithstanding that the sum of nucleotide lengths of the genes constitutes only 17% of the total 11,463 nucleotides of the original 13 genes. If only informative characters are considered, the share of the 3 genes is also 17%.

This observation and the fact that the third codon omission (equivalent to deletion of 33% of total characters and 55% of informative characters: Table 2) did not improve congruence among mammalian mitochondrial genes further confirm that the incongruence comes from specific gene sequences rather than from common partitions within genes. This prompted us to search for the origins of this anomalous phylogenetic signal of ND6, COII, and COIII.

It is known that the COII subunit of mitochondrial respiratory complex IV exhibits one of the most heterogeneous rates of amino acid substitution among placental mammals (Adkins *et al.*, 1996). There are isolated reports in the literature concerning the special phylogenetic properties of COII and COIII (Adkins *et al.*, 1996; Baker and DeSalle, 1997). Adkins and colleagues (1996) collected a large data set of COII sequences from several mammalian orders and, using phylogenetic hypotheses based on data independent of COII gene, demonstrated that an increased number of amino acid replacements are concentrated in this gene among higher primates. Baker and DeSalle (1997), while studying phylogeny of Hawaiian Drosophilids using mtDNA sequences, found that four of the genes including COII and COIII are significantly incongruent with the rest of the eight genes that they used.

The clue to these facts seems to be in functional roles that these subunits play in their respective electron transport complexes. Cytochrome oxidase (Complex IV) generates a transmembrane proton gradient by transferring electrons from cytochrome *c*, a water-soluble hemoprotein, to the active site, which contains a heme iron and a copper, and they are used to reduce O_2 into two water molecules (Grigorieff, 1999). The protons needed for this reaction are taken from the mitochondrial matrix side through two channels. Mammalian cytochrome oxidase contains 13 subunits. The 3 major subunits are encoded in mtDNA and form a functional core of the enzyme; this core is surrounded by 10 nuclear-encoded

small subunits. COI contains a low-spin heme (cytochrome *a*), the bimetallic cytochrome *a*3/CuB active site (Grigorieff, 1999) participating in the redox reaction. This key functional importance could be the reason that COI is highly conserved among vertebrates and presumably has accumulated less homoplasious changes than the other two mitochondria-encoded subunits. In contrast, the COII subunit contains only one of the Complex IV copper centers and the COIII subunit does not contain any redox centers. Moreover, the function of the COIII subunit has not been established yet, except for its binding to phospholipids (Grigorieff, 1999). A mutagenesis study (Haltia *et al.*, 1991) suggests that the COIII subunit does not participate in electron transfer reactions, but may play a role during biosynthesis of the multisubunit enzyme. Thus, the accessory or secondary role of the COII and COIII subunits could explain their lower degree of evolutionary conservation, which results in very low levels of congruence with the rest of the mitochondrially encoded proteins including COI.

As for NADH:ubiquinone oxidoreductase (Complex I), it is the largest and least understood of the mitochondrial enzymes. It is reasonable to hypothesize that the functional importance of ND2, ND4, and ND5 predicated a higher degree of conservation of their amino acid sequences and thereby a significant congruence with other mitochondrial protein-coding genes combined and presumed lower levels of accumulated homoplasy (Grigorieff, 1999). In contrast, the location and function of the other mitochondria-encoded sub-units, including ND6, is not known; however, there are some indications that the ND6 subunit mainly participates in the Complex I assembly (Bai and Attardi, 1998). Again, this ancillary role of ND6, as opposed to the primary functional importance of ND2, ND4, ND5, could be one of the reasons that the ND6 sequence is less conserved and more homoplasious. Interestingly, ND6 is also the only mitochondrial protein-coding gene encoded on the complementary strand with respect to the other

12 genes in higher vertebrates and it is the only gene that is incongruent with all other protein-coding genes (Table 4).

These data allow for the hypothesis that mitochondrially encoded subunits performing complementary or secondary functions in their respective enzymes, i.e., COII, COIII, and ND6, were less conserved and/or evolved with a different bias and are thus incongruent with the rest of the mitochondrial protein-coding genes. In contrast, subunits located in an active site of their enzyme complex and playing a key chemical role, e.g., COI, ND2, ND4, ND5, *ATP6*, and *ATP8* (Grandier-Vazeille *et al.*, 1994), are generally congruent with the rest of the matrix with or without weighting (Table 3).

We also investigated the phylogenies resulting from the deletion of ND6, COII, and COIII altogether. Heuristic searches yielded two different trees (a strict consensus of these is presented in Fig. 3). One of the trees had topology identical to that of the original tree containing all the genes equally weighted (Fig. 1A and data not shown). In the second tree (data not shown), Edentata, Proboscidea, and Tubulidentata were grouped basal to the Artiodactyla, Cetacea, Perrisodactyla, and Carnivora clade, with the rest of the taxa having the same positions. The increased number of most parsimonious trees suggests that the deletion of the three incongruent genes diminished resolution and thus phylogenetic signal (Cunningham, 1997a).

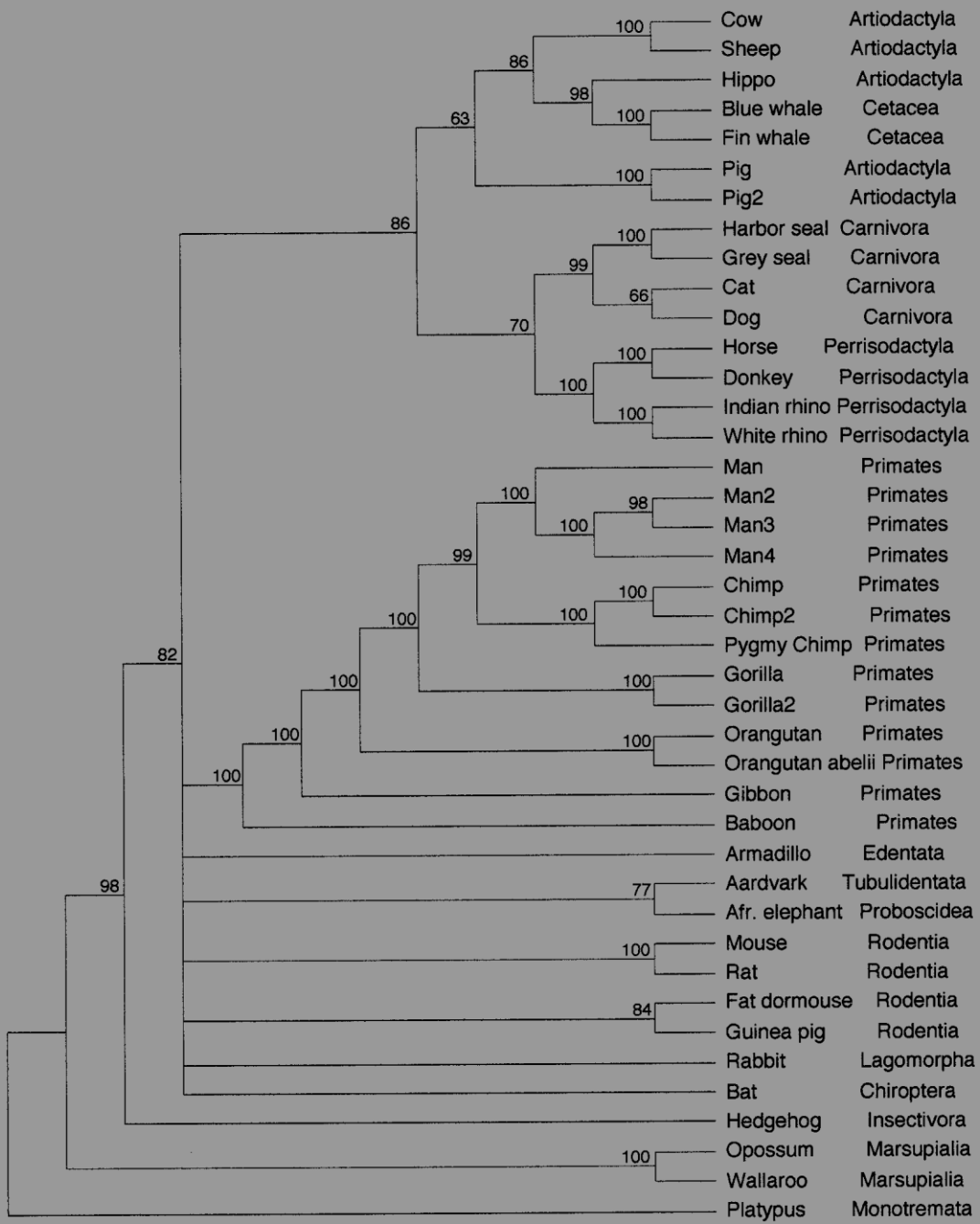

Cow Artiodactyla
Sheep Artiodactyla
Hippo Artiodactyla
Blue whale Cetacea
Fin whale Cetacea
Pig Artiodactyla
Pig2 Artiodactyla
Harbor seal Carnivora
Grey seal Carnivora
Cat Carnivora
Dog Carnivora
Horse Perrisodactyla
Donkey Perrisodactyla
Indian rhino Perrisodactyla
White rhino Perrisodactyla
Man Primates
Man2 Primates
Man3 Primates
Man4 Primates
Chimp Primates
Chimp2 Primates
Pygmy Chimp Primates
Gorilla Primates
Gorilla2 Primates
Orangutan Primates
Orangutan abelii Primates
Gibbon Primates
Baboon Primates
Armadillo Edentata
Aardvark Tubulidentata
Afr. elephant Proboscidea
Mouse Rodentia
Rat Rodentia
Fat dormouse Rodentia
Guinea pig Rodentia
Rabbit Lagomorpha
Bat Chiroptera
Hedgehog Insectivora
Opossum Marsupialia
Wallaroo Marsupialia
Platypus Monotremata

Fig. 3. Strict consensus of two most parsimonious trees obtained by simultaneous analysis of 10 mitochondrial genes (with COII, COIII, and ND6 excluded) with equal weighting of characters. Uninformative characters were excluded from analyses. Bootstrap scores are shown near the internal branches. Tree length, CI, and RI are listed in Table 2.

The decrease in resolution following attempts to improve congruence by weighting or sampling of characters calls into question the validity of such methods for the improvement of phylogenetic accuracy. This decrease in resolution suggests that the anomalous genes ND6, COII, and COIII along with homoplasious characters do contain some information necessary for unambiguous placement of taxa on a tree and that phylogenetic signal contained in these informative characters can overcome homoplasy carried in other parts of the genes. Thus, it might be tempting for some investigators to try to find stretches of DNA sequence that are informative and/or stretches that are homoplasious within genes and to try to improve congruence and reduce noise by editing, for example, by optimizing characters in a complete data set. This situation in reality may be more complicated if, for example, none of the sequences are completely homoplasious or completely informative and one needs to deal with individual characters. This makes such phylogenetic editing a daunting task.

A much simpler approach, however, can yield satisfactory results. As one can see here, equal treatment of all data can produce an unambiguous highly resolved phylogeny, identical to a tree produced with congruent data obtained by omitting evidence. Therefore, improvement of congruence in and of itself is not sufficient for improvement of phylogenetic accuracy, but rather increases in congruence can often lead to a drop in resolution. Since it is difficult to distinguish characters a priori in terms of their information content and homoplasy one by one, the best approach in phylogenetic reconstruction would be to analyze

all available characters as equally weighted. On the other hand, Cunningham (1997a), using six-parameter weighting in an iterative procedure, could show in some instances simultaneous improvement in congruence and resolution. However, Cunningham (1997a) used a much smaller data matrix, and his conclusions may hold true only for small matrices. In our case, weighting strategies did not significantly affect incongruence of the mitochondrial protein-coding genes, probably due to the sheer size of the matrix, which contains almost half a million characters. Therefore, if one seeks to improve resolution in a large data set by weighting, he or she may not be able to achieve improvement of congruence in the first place (Edwards *et al.*, 1997). As Allard *et al.* (1999a) and Cunningham (1997b) have shown, combination and analysis of incongruent data often will still produce strongly supported phylogeny, despite high levels of homoplasy in the data. Apparently, the signal contained in informative characters usually overcomes phylogenetic noise made by homoplasious characters [see Kallersjo *et al.*, (1999) and Wenzel and Siddall (1999) for a simulation of this phenomenon].

The special properties of the mammalian mitochondrial genome, e.g., maternal inheritance and the presence of orthologous single-copy genes, make the mitochondrial DNA a powerful tool for phylogenetic studies in higher vertebrates. We show here that conventional methods of reducing homoplasy such as weighting may not work well for large data sets. Since weighting strategies in this case may not yield greater congruence, but instead decrease resolution and/or increase the number of most parsimonious trees, different approaches (e.g., taxonomic or gene sampling) should be considered for the analysis of congruence in a large DNA sequence data set. Although homoplasious characters have accumulated to some extent in almost every gene, some mitochondrial genes (e.g., ND6, COII, and COIII), may have accumulated higher levels of ho-

moplasy due to the accessory nature of their function and may significantly affect congruence if introduced into a data set. As shown here, ND6, COII, and COIII, though highly incongruent with the rest of the molecular data, nevertheless carry the phylogenetic signal necessary for reconstruction of a highly resolved topology and therefore should be included in the analysis of a combined data set. It is known that morphological and fossil data are often incongruent with DNA sequence data with regard to cladistic analysis (Miyamoto, 1996). However, taken separately, neither type of data seems to be a source of stable phylogeny. For example, analysis of morphological data may effectively resolve one part of phylogeny, but yield a poor resolution or contradictory results in another part (O'Leary, 1999), where molecular data seem to perform better and vice versa in other clades. It would therefore be of interest in the future to try to reduce these major sources of phylogenetic information to a common denominator and analyze them in a combined data set (Allard *et al.*, 1999b). This strategy based on equal treatment of all characters and combined analysis of all available, albeit incongruent, data in our opinion will yield a better-resolved topology and ultimately should resolve many controversies concerning mammalian phylogeny.

ACKNOWLEDGMENTS

We thank Ginny Emerson, Marymegan Daly, and Krissy M. Kivimaki for assistance in editing the manuscript. The project was supported by NSF Grant DEB-9629319 to M. W. Allard. A Presidential Merit Fellowship in Biomedical Sciences supported N. A. Shevchuk.

REFERENCES

Adkins, R. M., Honeycutt R. L., and Disotell T. R. (1996). Evolution of eutherian cytochrome *c* oxidase subunit II: Heterogeneous rates of protein evolution and altered interaction with cytochrome *c*. *Mol. Biol. Evol.* 13: 1393–1404.

Allard, M. W., and Carpenter, J. M. (1996). On weighting and congruence. *Cladistics* 12: 183–198.

Allard, M. W., Farris, J. S., and Carpenter, J. (1999a). Congruence among mammalian mitochondrial genes. *Cladistics* 15: 75–84.

Allard, M. W., Honeycutt, R. L., and Novacek, M. J. (1999b). Advances in higher level mammalian relationships. *Cladistics* 15:
213–220.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., and Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–465.

Anderson, S., De Bruijn, M. H. L., Coulson, A. R., Eperon, E. C., Sanger, R., and Young, I. G. (1982). Complete sequence of bovine mitochondrial DNA: Conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* 156: 683–717.

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

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

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

Arnason, U., Gullberg, A., Johnsson, E., and Ledje, C. (1993). The nucleotide sequence of the mitochondrial DNA molecule of the grey seal, *Halichoerus grypus*, and a comparison with mitochondrial sequences of other true seals. *J. Mol. Evol.* 37: 323–330.

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

Arnason, U., Gullberg, A., and Xu, X. (1996b). 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., Gullberg, A., and Janke, A. (1997). Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between *Xenarthra* (Edentata) and Ferungulates. *Mol. Biol. Evol.* 14: 762–768.

Arnason, U., Gullberg, A., and Janke, A. (1998). Molecular timing of primate divergences as estimated by two nonprimate calibration points. *J. Mol. Evol.* 47: 718–727.

Arnason, U., Gullberg, A., and Janke, A. (1999). The mitochondrial DNA molecular phylogeny 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.

Bai, Y., and Attardi, G. (1998). The mtDNA-encoded ND6 subunit of mitochondrial NADH dehydrogenase is essential for the assembly of the membrane arm and the respiratory function of the enzyme. *EMBO J.* 17: 4848–4858.

Baker, R. H., and DeSalle, R. (1997). Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst. Biol.* 46: 654–673.

Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W., and Clayton, D. A. (1981). Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26: 167–180.

Cunningham, C. W. (1997a). Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Syst. Biol.* 46: 464–478.

Cunningham, C. W. (1997b). Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol.* 14: 733–740.

D'Erchia, A. M., Gissi, C., Pesole, G., Saccone, C., and Arnason, U. (1996). The guinea-pig is not a rodent. *Nature* 381: 597–600.

DeSalle, R., and Brower, A. V. Z. (1997). Process partitions, congruence, and the independence of characters: Inferring relationships among closely related Hawaiian *Drosophila* from multiple gene regions. *Syst. Biol.* 46: 751–764.

Edwards, S. V., Chesnut, K., Satta, Y., and Wakeland, E. K. (1997). Ancestral polymorphism of Mhc class II genes in mice: Implications for balancing selection and the mammalian molecular clock. *Genetics* 146: 655–668.

Farris, J. S. (1989). The retention index and the rescaled consistency index. *Cladistics* 5: 417–419.

Farris, J. S., Kallersjo, M., Kluge, A. G., and Bult, C. (1994). Testing significance of incongruence. *Cladistics* 10: 315–319.

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

Gissi, C., Gullberg, A., and Arnason, U. (1998). The complete mitochondrial DNA sequence of the rabbit, *Oryctolagus cuniculus*. *Genomics* 50: 161–169.

Grandier-Vazeille, X., Ouhabi, R., and Guerin, M. (1994). Antibodies against subunits of F0 sector of ATP synthase from *Saccharomyces cerevisiae*: Stimulation of ATP synthase by subunit-8-reactive antibodies and inhibition by subunit-9-reactive antibodies. *Eur. J. Biochem.* 223: 521–528.

Grigorieff, N. (1999). Structure of the respiratory NADH:ubiquinone oxidoreductase (complex I). *Curr. Opin. Struct. Biol.* 9: 476–483.

Haltia, T., Saraste, M., and Wikstrom, M. (1991). Subunit III of cytochrome *c* oxidase is not involved in proton translocation: A site-directed mutagenesis study. *EMBO J.* 10: 2015–2021. Hiendleder, S., Mainz, K., Plante, Y., and Lewalski, H. (1998). Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: No evidence for contributions from urial and argali sheep. *J. Hered.* 89: 113–120.

Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S., and Takahata, N. (1992). Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* 35: 32–43.

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

Janke, A., Feldmaier-Fuchs, G., Thomas, W. K., Von Haeseler, A., and Pa¨a¨bo, S. (1994). The marsupial mitochondrial genome and the evolution of placental mammals. *Genetics* 137: 243–256.

Janke, A., Gemmell, N. J., Feldmaier-Fuchs, G., von Haeseler, A., and Pa¨a¨bo, S. (1996). The mitochondrial genome of a monotreme— the platypus (*Ornithorhynchus anatinus*). *J. Mol. Evol.* 42: 153–159.

Janke, A., Xu, X., and Arnason, U. (1997). The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationships among Monotremata, Marsupialia, and Eutheria. *Proc. Natl. Acad. Sci. USA* 94: 1276–1281.

Kallersjo, M., Albert, V. A., and Farris, J. S. (1999). Homoplasy increases phylogenetic structure. *Cladistics* 15: 91–94.

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

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

Lecointre, G., Rachdi, L., Darlu, P., and Denamur, E. (1998). *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol. Biol. Evol.* 15: 1685–1695.

Lin, C. S., Sun, Y. L., Liu, C. Y., Yang, P. C., Chang, L. C., Cheng, I. C., Mao, S. J. T., and Huang, M. C. (1999). Complete nucleotide sequence of pig (*Sus scrofa*) mitochondrial genome and dating evolutionary divergence within Artiodactyla. *Gene* 236: 107–114.

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

Miyamoto, M. M. (1996). A congruence study of molecular and morphological data for eutherian mammals. *Mol. Phylogenet. Evol.* 6: 373–390.

Miyamoto, M. M., and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* 44: 64–76.

Montoya, J., Ojala, D., and Attardi, G. (1981). Distinctive features of the 59-terminal sequences of the human mitochondrial mRNAs. *Nature* 290: 465–470.

Naylor, G. J. P., and Brown, W. M. (1998). Structural biology and phylogenetic estimation. *Nature* 388: 527–528.

Nixon, K. C. (2000). Winclada (beta) ver. 0.9.9. Published by the author. Ithaca, NY. [Available at http://www.cladistics.com].

Novacek, M. J. (1992). Fossils, topologies, missing data, and the higher level phylogeny of eutherian mammals. *Syst. Biol.* 41: 58–73.

O'Leary, M. (1999). Parsimony analysis of total evidence from extinct and extant taxa and the cetacean–artiodactyl question (Mammalia, Ungulata). *Cladistics* 15: 315–330.

Pumo, D. E., Finamore, P. S., Franek, W. R., Phillips, C. J., Tarzami, S., and Balzarano, D. (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.

Remsen, J., and DeSalle, R. (1998). Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. *Mol. Phylogenet. Evol.* 9: 225–235.

Reyes, A., Pesole, G., and Saccone, C. (1998). Complete mitochondrial DNA sequence of the fat dormouse, *Glis glis*: Further evidence of rodent paraphyly. *Mol. Biol. Evol.* 15: 499–505.

Swofford, D. L. (2000). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0a2b. Sinauer, Sunderland, MA.

Thornton, J. M., and DeSalle, R. (2000). A method to localize and test the significance of incongruence: Detecting domain shuffling in the nuclear receptor superfamily. *Syst. Biol.* 49: 183–201.

Ursing, B. M., and Arnason, U. (1998a). Analyses of mitochondrial genomes strongly support a hippopotamus–whale clade. *Proc. R. Soc. Lond. B Biol. Sci.* 265: 2251–2255.

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

Wenzel, J. W., and Siddall, M. E. (1999). "Noise." *Cladistics* 15: 51–64.

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

Xu, X., and Arnason, U. (1996a). The mitochondrial DNA molecule of the Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. *J. Mol. Evol.* 43: 431–437.

Xu, X., and Arnason, U. (1996b). A complete sequence of the mitochondrial genome of the Western lowland gorilla. *Mol. Biol. Evol.* 3: 691–698.

Xu, X., Gullberg, A., and Arnason, U. (1996a). The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *J. Mol. Evol.* 43: 438–446.

Xu, X., Janke, A., and Arnason, U. (1996b). The complete mitochondrial DNA sequence of the greater Indian rhinoceros, *Rhinoceros nicornis*, and the phylogenetic relationships among the Carnivora, Perissodactyla, and Artiodactyla (lCetacea). *Mol. Biol. Evol.* 13: 1167–1173.

Xu, X., and Arnason, U. (1997). The complete mitochondrial DNA sequence of the white Rhinoceros, *Ceratotherium simum*, and comparison with the

mtDNA sequences of the Indian Rhinoceros, *Rhinceros unicornis. Mol. Phylogenet. Evol.* 7: 189–194.