ISRAEL JOURNAL OF ZOOLOGY, Vol. 47, 2001, pp. 259–270

# USAGE OPTIMIZATION OF UNEVENLY SAMPLED DATA THROUGH THE COMBINATION OF QUARTET TREES: A EUTHERIAN DRAFT PHYLOGENY BASED ON 640 NUCLEAR AND MITOCHONDRIAL PROTEINS

MARC ROBINSON-RECHAVI[a,*] AND DAN GRAUR[b]

[a]*Laboratoire de Biologie Moléculaire et Cellulaire, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon cedex 07, France*

[b]*Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel*

## ABSTRACT

Molecular phylogeneticists must frequently decide on a painful trade-off between the number of taxa and the number of sequences used in a study. Here, we illustrate the advantages of a method of combining quartet trees to solve this dilemma. We apply the method to a data set of 640 protein-sequence alignments from 4 to 24 eutherian taxa, and obtain a global eutherian phylogeny. In agreement with recent studies, we identify three major super-ordinal clades. The first clade is Afrotheria, a cluster of endemic African mammals. The second clade is an emended Laurasiatheria, consisting of Cetartiodactyla (cetaceans, ruminants, hippopotamuses, pigs, and tylopods), Perissodactyla (horses and rhinoceroses), Carnivora, Pholidota (pangolins), Chiroptera (bats), and Erinaceidae (hedgehogs). A tentatively identified third clade consists of some archontans (primates, flying lemurs, and tree shrews) as well as lagomorphs and rodents. Evolutionary relations within these major clades are well resolved. We also show that nuclear encoded proteins resolve eutherian phylogeny better than complete mitochondria. Finally, our results demonstrate that combining quartet trees provides a major opportunity to resolve unevenly sampled complex phylogenies.

## INTRODUCTION

Because molecular data are not available evenly for all taxa of interest, phylogeneticists must frequently decide on a painful trade-off between the number of taxa and the number of sequences used in a study. If we restrict ourselves only to sequences that are available for all the taxa under study, we may end up ignoring the vast majority of sequences. If, on the other hand, we increase the number of sequences, we end up seriously restricting taxonomic representation. Here, we take advantage of a method of combining quartet trees (Ben-Dor et al., 1998) to solve this dilemma, and obtain a unique eutherian phylogeny from 640 protein sequences.

*Author to whom correspondence should be addressed. E-mail: marc.robinson@ens-lyon.fr

Accepted November 2001.

Despite much attention from systematists and paleontologists, the phylogenetic relationships among eutherian orders have remained highly controversial (Novacek, 1992; Shoshani and McKenna, 1998). Among major successes of molecular phylogeny (De Jong, 1998) are the solid nesting of Cetacea inside Artiodactyla, and the definition of a new super-ordinal clade, Afrotheria. Molecular studies have also prompted heated debates, notably on the monophyly of rodents or the position of Lagomorpha. By combining 640 protein trees, we build a phylogeny that takes into account proteins known in many taxa, such as the von Willebrand factor known in all of the 24 taxa used in this study, as well as 258 orthologous proteins whose sequences are known in primates, murids, ruminants, and an outgroup taxon. It also allows us to compare phylogenetic degrees of resolution between trees obtained with mitochondrial data and those obtained with nuclear data.

## MATERIALS AND METHODS

Twenty-three eutherian taxa were predefined (Table 1) under the assumption that each taxon is monophyletic. Most of these taxa correspond to traditional orders, but we split three orders for which there is a strong suggestion of paraphyly (Graur and Higgins, 1994; Stanhope et al., 1998; Nikaido et al., 1999). Artiodactyla was divided into Hippopotamidae, Ruminantia, Suina (= Suoidea), and Tylopoda. Lipotyphla (= Insectivora) was separated into Chrysochloridae (= Afrosoricida) and Erinaceidae (= Erinaceomorpha); other sub-orders were not considered due to paucity of data. In Rodentia, we only sampled two lineages (Muridae and Hystricognathi). The twenty-fourth taxon consisted of outgroups.

For all protein-coding genes known in at least four of the 23 predefined eutherian taxa, or 3 and an outgroup, one protein sequence per taxon was extracted from release 27 of HOVERGEN (Duret et al., 1994). This data set was complemented with sequences from release 34 of SWISS-PROT + TrEMBL (Bairoch and Apweiler, 1999) and release 55 of PIR (Barker et al., 1999), and updated for the 15 most poorly sampled taxa (Table 1), as well as for mammalian outgroups from release 40 of HOVERGEN. In addition, proteins were deduced from complete mitochondrial sequences obtained through the NCBI web site, to which recently sequenced genes were added (Liu et al., 2001; Madsen et al., 2001; Murphy et al., 2001). All aligned sequences are deposited at http://www.ens-lyon.fr/~mrobinso/mammals.dat.

Paralogy was excluded by individual examination of all gene sets, taking into account visible duplications, evolutionary rates, very unusual phylogenetic topologies, sequence annotations, and information from the primary literature. When several orthologous sequences were available within a higher taxon, the sequence with the shortest branch length in a neighbor-joining tree was used, to minimize long branch attraction artifacts. Outgroups were always orthologous sequences from amniotes, with preference given to the shortest branch. Sequences were aligned by ClustalW (Thompson et al., 1994), and all alignments were checked manually using SEAVIEW (Galtier et al., 1996). The final data set consists of 640 sets of aligned proteins, each set consisting of sequences from 4 to 24 taxa.

Table 1
Distribution of protein data among taxa

| Taxon | Number of protein sequences | | |
|---|---|---|---|
| | Nuclear | Mitochondrial | Total |
| Primates | 624 | 13 | 637 |
| Muridae | 612 | 13 | 625 |
| Ruminantia | 443 | 13 | 456 |
| Suina | 341 | 13 | 354 |
| Lagomorpha | 281 | 13 | 294 |
| Carnivora | 211 | 13 | 224 |
| Hystricognathi | 126 | 13 | 139 |
| Perissodactyla | 95 | 13 | 108 |
| Cetacea | 31 | 13 | 44 |
| Scandentia | 25 | 13 | 38 |
| Erinaceidae | 18 | 13 | 31 |
| Proboscidea | 20 | 13 | 33 |
| Chiroptera | 19 | 13 | 32 |
| Hippopotamidae | 20 | 13 | 33 |
| Xenartha | 18 | 13 | 31 |
| Tylopoda | 28 | 1 | 29 |
| Tubulidentata | 16 | 13 | 29 |
| Macroscelidea | 15 | 13 | 28 |
| Sirenia | 17 | 1 | 18 |
| Hyracoidea | 14 | 1 | 15 |
| Dermoptera | 16 | 1 | 17 |
| Pholidota | 14 | 1 | 15 |
| Chrysochloridae | 3 | 1 | 4 |
| Outgroup | 358 | 13 | 371 |
| *Metatheria+Prototheria* | 57 | 13 | 70 |
| *Sauropsida* | 301 | 0 | 301 |

For each gene, a tree was built by maximum likelihood or distance methods. Quartet puzzling (Strimmer and Von Haeseler, 1996) was used for maximum likelihood, allowing reasonable time estimation of support values, with the JTT substitution model (Jones et al., 1992) for nuclear genes and the mtREV24 model (Adachi and Hasegawa, 1996b) for mitochondrial genes, and gamma-distributed rate heterogeneity (as estimated by TREE-PUZZLE for each protein, with 8 categories). For distances, BIONJ (Gascuel, 1997), an improved neighbor-joining algorithm, was applied to Poisson-corrected distances. For 72 genes BIONJ was not computable, and a classical neighbor-joining algorithm (Saitou and Nei, 1987) was used. Support for nodes in BIONJ or NJ trees was estimated by 1,000 bootstrap replicates. Whatever the tree-building method, for each set of four taxa and each gene, a quartet tree was deduced from the gene tree, separating two taxa from the other two. Each quartet tree was weighted by a score defined as the highest support value (bootstrap or quartet puzzling) separating the four taxa in the gene tree; in

other words, by the support for this quartet in the tree. The quartet trees thus obtained were then combined using AllTree (Ben-Dor et al., 1998). First all subsets of the taxa are computed and the division which satisfied the most quartet trees is retained. Then an exact search is done on each subset, yielding the overall tree satisfying the most quartet trees.

Reliability of nodes in the total tree was assessed by counting the proportion of relevant quartets (deduced from gene trees) that supported each node. This procedure was repeated by using only quartet trees with more than 90% support.

## RESULTS AND DISCUSSION

We have analyzed a total of 202,170 aligned and ungapped sites, with an average of 316 sites per protein. The number of taxa per protein varies from 4 to 24, with a mean of 5.63 taxa per protein (Table 2). For comparison, in a complete mitochondria analysis, 3669 complete sites are available for 18 out of the 24 taxa we analyzed.

Because of the novelty of the method used, two points should be made: (i) the support indices on the internal branches (Figs. 1,2,3B) are not bootstrap values, but proportions of genes that support a given branch. They do not represent a statistical test, but rather a summary of the information we have gathered from different genes. And (ii) we did not systematically build four-taxa ("quartet") trees from raw protein data, but rather decomposed each gene tree of *n* taxa into all its constituent quartets, as a way of representing the phylogenetic information in a manner that allows easier combination of information from different trees. Notwithstanding, each gene tree was built with as many taxa as possible, within the limits of our choice of one sequence per pre-defined taxon. Thus, the use of gene trees to build the global tree (see Methods) avoids possible problems associated with quartet methods (Adachi and Hasegawa, 1996a).

Whereas some nodes are supported by less than half of all relevant gene trees (Fig. 1), all nodes are supported by a majority of relevant branchings with support above 90% (stars in Fig. 1). Moreover, all nodes are supported by a majority of distance-based gene trees, although the global tree was built from quartet puzzling likelihood trees. And there is an excellent correlation between support of nodes by both tree-building methods. These observations are encouraging, indicating a high level of consistency of the methodology used.

Table 2

Taxonomic sampling of proteins

| Number of taxa | Number of proteins |
|---|---|
| 4 | 319 |
| 5 | 138 |
| 6 | 86 |
| 7 | 43 |
| ≥ 8 | 54 |

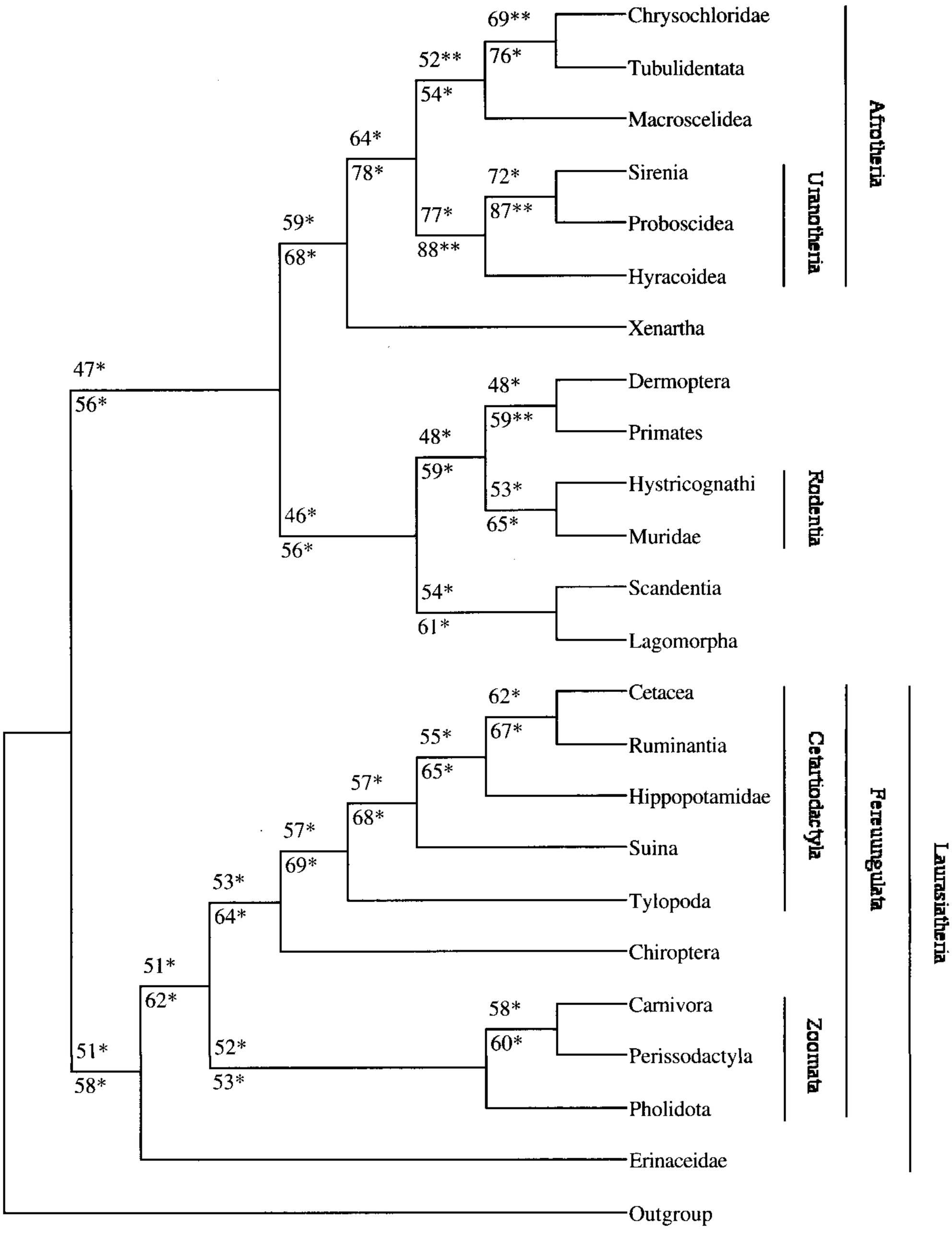



Fig. 1. Phylogenetic tree of eutherian mammals. Global phylogenetic tree obtained by combining the quartet trees deduced from 640 quartet puzzling maximum-likelihood gene trees. Branch lengths are arbitrary. Numbers above branches are support indices from maximum-likelihood gene trees. Numbers under branches are support indices from distance gene trees. Support is the proportion (in %) of relevant quartet trees (deduced from gene trees) that support a given branch, and it should not be confused with bootstrap values. * indicates a branch supported by more than half of relevant quartets with bootstrap over 90%. ** indicates a branch supported by all relevant quartets with bootstrap over 90%.

*SUPER-ORDINAL EUTHERIAN CLADES*

The overall topology of the eutherian tree (Fig. 1) indicates three basal or near-basal clades: an emended Laurasiatheria (Waddell et al., 1999) (clade IV of Murphy et al., 2001), Afrotheria (Stanhope et al., 1998) (clade I of Murphy et al., 2001), and an emended Archontan clade (clade III of Murphy et al., 2001) which groups with Afrotheria. This overall topology is consistent with the findings of several recent studies with large taxonomic sampling (Liu et al., 2001; Madsen et al., 2001; Murphy et al., 2001).

Of these three major clades, the one that receives the highest support is Afrotheria, a grouping of endemic African mammals (Stanhope et al., 1998; Springer et al., 1999). This clade and the branchings inside it are supported by the highest proportions of quartets, often including all trees with bootstrap scores over 90% (Fig. 1). Recent studies are contradictory concerning relations among afrotherians. The relations we recover with all data are in agreement with the molecular and morphological data of Liu et al. (Liu et al., 2001). Our position of Macroscelidea is in disagreement with the molecular study of Murphy et al. (2001), whereas the same branching order inside Uranotheria is recovered. When only nuclear genes are used in our study (Fig. 2A), the branching order inside Uranotheria changes, in agreement with that proposed by Madsen et al. (2001), but support is lower than in the global tree.

As evident from Fig. 1, our method recovers Fereuungulata, a grouping of Carnivora, Cetartiodactyla (Cetacea + Artiodactyla), Perissodactyla, and Pholidota (Waddell et al., 1999). We also recover Laurasiatheria (Fereuungulata + Chiroptera + Erinaceidae) and Cetartiodactyla (Graur and Higgins, 1994; Shimamura et al., 1997). However, with nuclear + mitochondrial data we do not recover the sister relationship between Cetacea and Hippopotamidae (Whippomorpha) (Nikaido et al., 1999). We note that nuclear data (Fig. 2A) yields identical relationships among members of Cetartiodactyla as those recovered from data on interspersed elements (Nikaido et al., 1999). Cetartiodactyla is not recovered by our mitochondrial data (Fig. 2B). The grouping of Perissodactyla and Carnivora (and Pholidota), recovered in all our analyses, was suggested on the basis of complete mitochondrial data (Xu et al., 1996), but an analysis of 36 concatenated nuclear and mitochondrial proteins (Graur et al., 1997) gave weak support for a grouping of Perissodactyla and Artiodactyla instead, and recent studies do not support our grouping (Liu et al., 2001; Madsen et al., 2001; Murphy et al., 2001). It should be noted that the protein data used in all of these studies is incorporated in our analysis. Thus we see that analysis of all data resolves the fereuungulate trichotomy by the grouping of Carnivora with Perissodactyla (and Pholidota) to the exclusion of Cetartiodactyla, but with low support values (Fig. 1).

The third large clade within Eutheria groups various orders whose positions are often debated (Allard et al., 1996). We observe a sister relationship between Dermoptera and Primates supported by all our analyses (Figs. 1 and 2), which in turn group with rodents in the global tree (Fig. 1). With either the nuclear proteins or with all the proteins, Dermoptera and Primates form a larger clade with Scandentia and Lagomorpha (Fig. 2A), consistent with other studies (Liu et al., 2001; Madsen et al., 2001; Murphy et al., 2001). This clade, which resembles Archonta, is as yet unnamed.

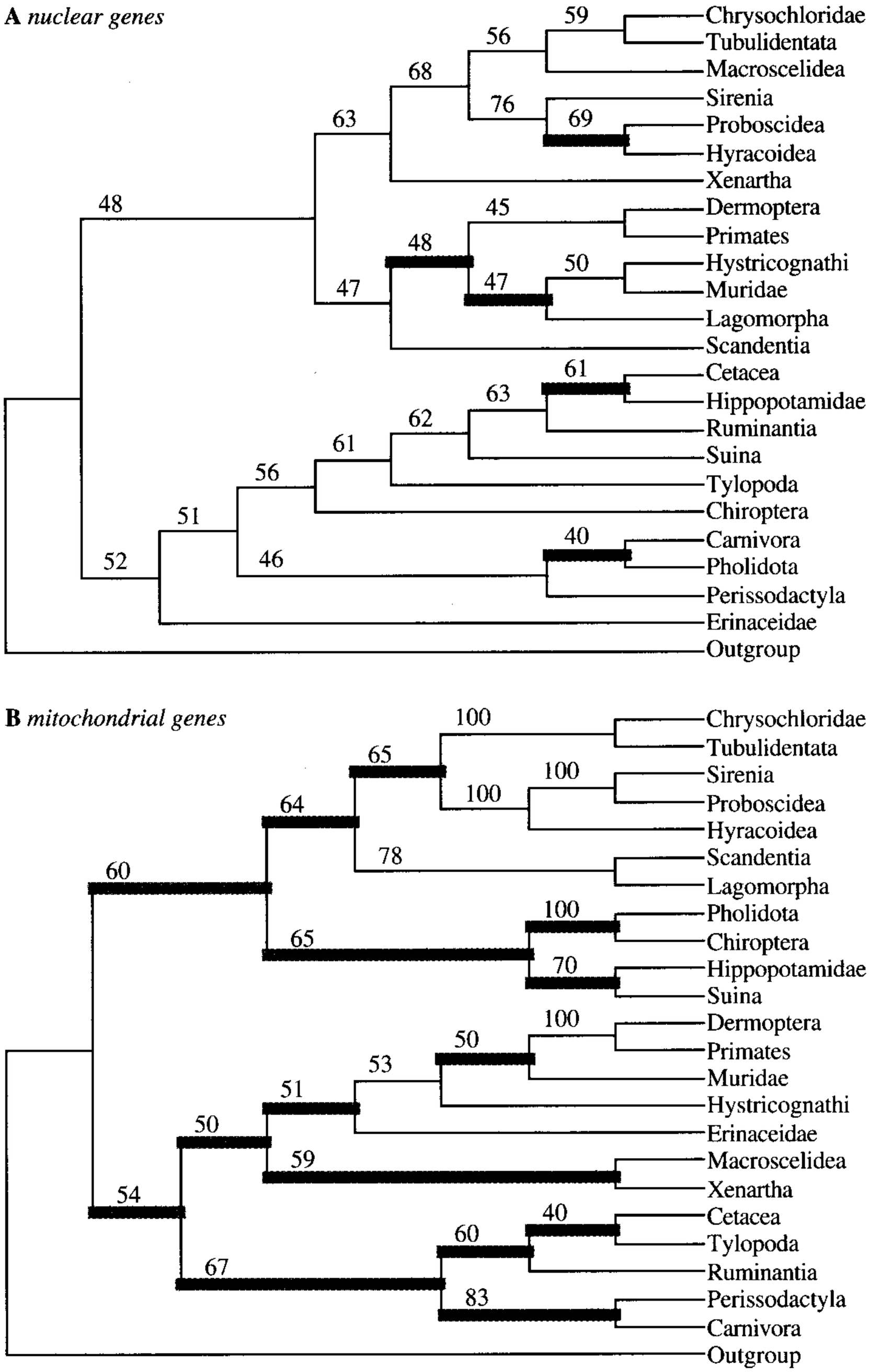



Fig. 2. Comparison of phylogenetic trees from nuclear and mitochondrial data. Phylogenetic trees obtained by combining the quartet trees deduced from quartet puzzling maximum-likelihood gene trees, for 627 nuclear-encoded genes (A) or for 13 mitochondrion-encoded genes (B). Branch lengths are arbitrary. Numbers above branches are support indices from maximum-likelihood puzzling gene trees. Support is the proportion (in %) of relevant quartet trees (deduced from gene trees) that support a given branch. Bold gray branches are not recovered in the combined nuclear + mitochondrial dataset (Fig. 1).

*POSITIONS OF RODENTS AND RABBITS*

The position of rabbits and allies (Lagomorpha), and notably their grouping with rodents within Glires or with rodents and macroscelids within Anagalida, has been extensively debated for many years. Glires and Anagalida have been supported by cladistic morphological analysis (Luckett and Hartenberger, 1993; McKenna and Bell, 1997), but rejected by molecular data (Graur et al., 1996; Gissi et al., 1998). Glires was recovered by recent molecular work with large taxonomic sampling (Huchon et al., 1999; Liu et al., 2001; Madsen et al., 2001; Murphy et al., 2001). The grouping of Lagomorpha with Scandentia, suggested by a quartet study of concatenated proteins (Graur et al., 1996) and by analysis of complete mitochondria (Schmitz et al., 2000), is supported in our global tree and with mitochondrial data only (Figs. 1, 2B). But with nuclear data only, Glires is recovered, albeit with very low support (Fig. 2A). We conclude that the inclusion of Lagomorpha within a clade including rodents, primates, and Scandentia is well-supported by protein sequences, but that its position inside the clade remains an open question.

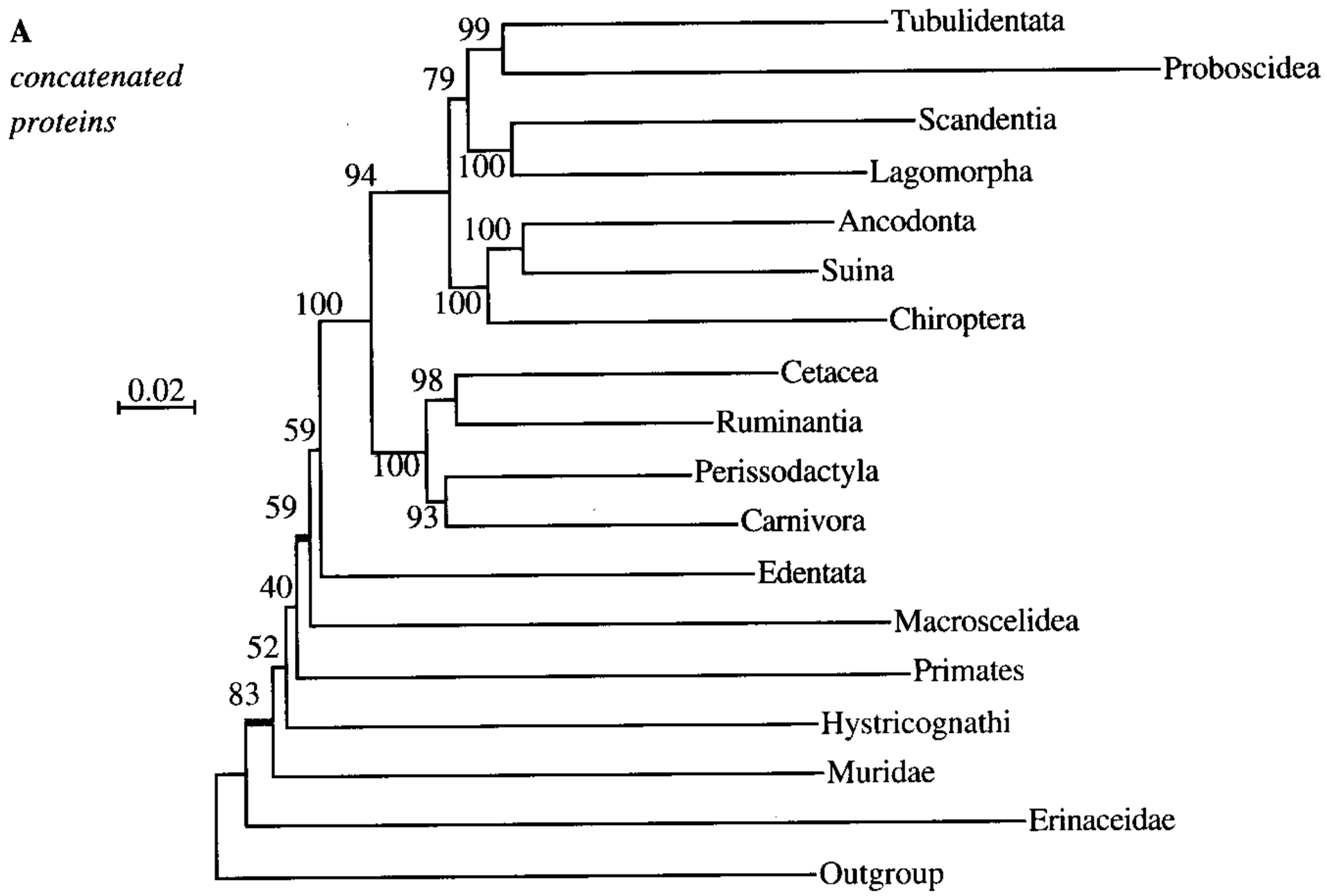
The monophyly of Rodentia has been challenged by molecular analyses (Graur et al., 1991; D'Erchia et al., 1996), whereas morphological data strongly support this clade (Luckett and Hartenberger, 1985, 1993). It has been recovered by recent molecular studies with large taxonomic sampling (Huchon et al., 1999; Robinson-Rechavi et al., 2000; Liu et al., 2001; Madsen et al., 2001; Murphy et al., 2001). Here murids and hystricognaths are grouped, except by mitochondrial data alone (Fig. 2B). Our results indicate that, like Lagomorpha, rodent lineages belong to an emended Archonta. The internal resolution inside this tentative clade is problematic.

*IS XENARTHA BASAL?*

Morphological analyses usually position Xenartha (= Edentata) as a basal branch, thus clustering all other eutherian orders into a clade called Epitheria (McKenna and Bell, 1997; Liu et al., 2001). In trees based on complete mitochondrial sequences, part of Lipotyphla (hedgehogs) emerges as the basal clade (Krettek et al., 1995), while Xenartha groups with Fereuungulata (Arnason et al., 1997). However, these molecular trees lack information on afrotherians. In our tree (Fig. 1) Xenartha groups with Afrotheria, in agreement with the molecular analysis of Madsen et al. (2001), but not with that of Murphy et al. (2001). Xenartha + Afrotheria form a larger clade with the emended Archonta in our analysis. Thus, with all available protein data, we do not find support for any order forming an early offshoot at the base of the eutherian tree.

*NUCLEAR VS. MITOCHONDRIAL DATA*

One of the main deficiencies of molecular phylogeny as far as eutherian mammals are concerned has been the contradictory results obtained when using complete mitochondrial data versus nuclear data. Comparison of the nuclear and mitochondrial trees (Fig. 2) with expectations based on morphology, interspersed elements, and other sources, tends to indicate that nuclear protein data are phylogenetically more reliable than mitochondrial data, at least as far as our taxonomic sampling is concerned. Notably, nuclear data recover Whippomorpha, Rodentia, Glires, and Afrotheria (Fig. 2A). In



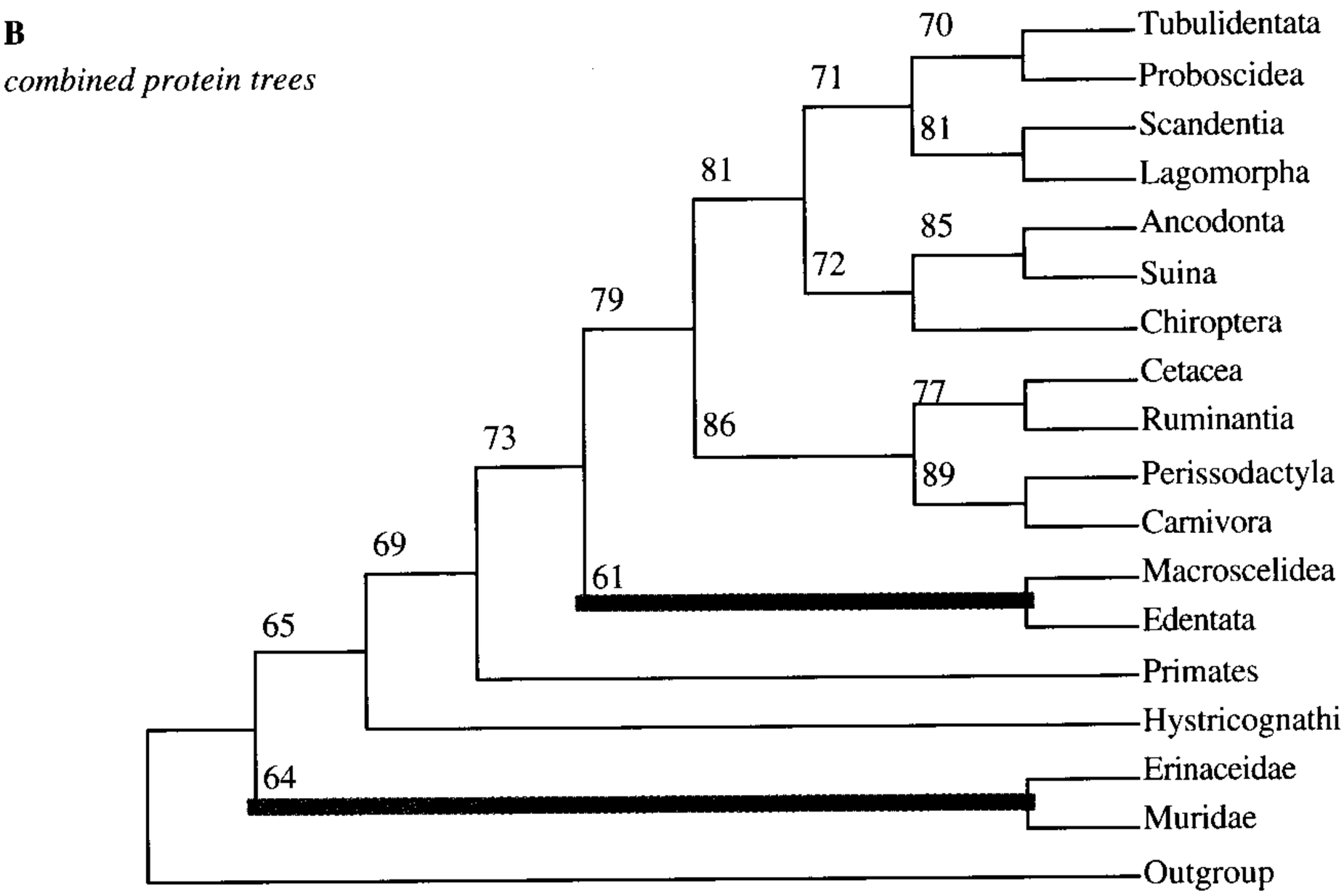




Fig. 3. Phylogenetic trees from complete mitochondria. Phylogenetic trees obtained using only mitochondrion-encoded proteins from taxa for which a complete mitochondrial sequence is known. (A) BIONJ tree obtained from all proteins concatenated in one "super-protein"; figures at nodes are bootstrap scores from 2000 bootstrap replicates (in %). (B) Tree obtained combining the quartet trees deduced from BIONJ protein trees; figures at nodes are the proportion (in %) of relevant quartet trees (deduced from gene trees) that support a given branch. (A and B) Bold gray branches differ between the two trees.

contrast, the mitochondrial data in this study, whether analyzed with "classical" methods (Fig. 3A) or by combining trees (Figs. 2B and 3B), recovers none of these clades, and generally has a low resolution for higher eutherian phylogeny. These observations are consistent with the conclusions of Springer et al. (2001), who show that nuclear encoded genes have a better resolution per residue than mitochondrial genes in mammalian phylogeny.

Using only proteins from complete mitochondria allows comparison between a classical approach, deducing the tree from concatenated proteins though BIONJ (Fig. 3A), or combining protein trees by AllTree (Fig. 3B). It is extremely encouraging that the two trees are mostly consistent, and also that the only two differences (in bold gray in Fig. 3B) correspond to the lowest support indices of the combined protein tree. Although we cannot provide an absolute cut-off value for this index, this validates it as correctly indicating support of nodes by the data.

## CONCLUSIONS

Our results show that combining gene trees can be a powerful tool for resolving complex phylogenies in cases where the sampling of genes is uneven among the constituent taxa. It also confirms that nuclear data outperform mitochondrial data, despite the inherent uneven sampling of the former relative to that of the latter. Finally, it appears that molecular results are converging towards a complete resolution of mammalian phylogeny, with no explosive radiation of eutherian orders at the base of the tree.

## ACKNOWLEDGMENTS

The initial data compilation and analyses were done at Tel Aviv University, while M.R.-R. received a Lavoisier grant from the French Ministry of Foreign Affairs. Additional support was provided by the Magnet Da'at Consortium of the Israeli Ministry of Industry and Trade. Complete mitochondrial sequences were provided before publication by Ú. Árnason, M. Hasegawa, and C. Saccone. We thank Laurent Duret, Manolo Gouy, Dorothée Huchon, Jean-Jacques Jaeger, Giddy Landan, Vincent Laudet, Dan Pelleg, and Tal Pupko for help and advice.

## REFERENCES

Adachi, J., Hasegawa, M. 1996a. Instability of of quartet analyses of molecular sequence data by the maximum likelihood method: the Cetacea/Artiodactyla relationships. Mol. Phylogenet. Evol. 6: 72–76.

Adachi, J., Hasegawa, M. 1996b. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. 42: 459–468.

Allard, M.W., Mcniff, B.E., Miyamoto, M.M. 1996. Support for interordinal Eutherian relationships with an emphasis on primates and their Archontan relatives. Mol. Phyl. Evol. 5: 78–88.

Árnason, Ú., Gullberg, A., Janke, A. 1997. Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenartha (Edentata) and ferungulates. Mol. Biol. Evol. 14:

762–768.
Bairoch, A., Apweiler, R. 1999. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. Nucleic Acids Res. 27: 49–54.
Barker, W.C., Garavelli, J.S., Mcgarvey, P.B. et al. 1999. The PIR-International protein sequence database. Nucleic Acids Res. 27: 39–43.
Ben-Dor, A., Chor, B., Graur, D., Ophir, R., Pelleg, D. 1998. Constructing phylogenies from quartets: elucidation of eutherian superordinal relationships. J. Comput. Biol. 5: 377–390.
D'Erchia, A.M., Gissi, C., Pesole, G., Saccone, C., Árnasson, Ú. 1996. The guinea pig is not a rodent. Nature 381: 597–600.
De Jong, W.W. 1998. Molecules remodel the mammalian tree. Trends Ecol. Evol. 13: 270–275.
Duret, L., Mouchiroud, D., Gouy, M. 1994. HOVERGEN: a database of homologous vertebrate genes. Nucleic Acids Res. 22: 2360–2365.
Galtier, N., Gouy, M., Gautier, C. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12: 543–548.
Gascuel, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. 14: 685–695.
Gissi, C., Gullberg, A., Árnason, Ú. 1998. The complete mitochondrial DNA sequence of the rabbit, *Oryctolagus cuniculus*. Genomics 50: 161–169.
Graur, D., Higgins, D.G. 1994. Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. Mol. Biol. Evol. 11: 357–364.
Graur, D., Hide, A.H., Li, W.H. 1991. Is the guinea-pig a rodent? Nature 351: 649–652.
Graur, D., Duret, L., Gouy, M. 1996. Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). Nature 379: 333–335.
Graur, D., Gouy, M., Duret, L. 1997. Evolutionary affinities of the order Perissodactyla and the phylogenetic status of the superordinal taxa Ungulata and Altungulata. Mol. Phylogenet. Evol. 7: 195–200.
Huchon, D., Catzeflis, F.M., Douzery, E.J. 1999. Molecular evolution of the nuclear von Willebrand Factor gene in mammals and the phylogeny of rodents. Mol. Biol. Evol. 16: 577–589.
Jones, D.T., Taylor, W.R., Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8: 275–282.
Krettek, A., Gullberg, A., Arnason, U. 1995. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphla. J. Mol. Evol. 41: 952–957.
Liu, F.-G.R., Miyamoto, M.M., Freire, N.P., Ong, P.Q., Tennant, M.R., Young, T.S., Gugel, K.F. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. Science 291: 1786–1742.
Luckett, W.P., Hartenberger, J.L. 1985. Evolutionary relationships among rodents: comments and conclusions. In: Luckett, W., Hartenberger, J., eds. Evolutionary relationships among rodents, a multidisplinary analysis. Plenum Press, New York, pp. 685–712.
Luckett, W.P., Hartenberger, J.L. 1993. Monophyly or polyphyly of the order Rodentia : possible conflict between morphological and molecular interpretations. J. Mammal. Evol. 1: 127–147.
Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., de Jong, W.W., Springer, M.S. 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409: 610–614.
McKenna, M.C., Bell, S.K. 1997. Classification of mammals above the species level. Columbia University Press, New York, 631 pp.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J. 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409: 614–618.

Nikaido, M., Rooney, A.P., Okada, N. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA 96: 10261–10266.

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

Robinson-Rechavi, M., Ponger, L., Mouchiroud, D. 2000. Nuclear gene LCAT supports rodent monophyly. Mol. Biol. Evol. 17: 1410–1412.

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

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

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., Okada, N. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388: 666–670.

Shoshani, J., McKenna, M.C. 1998. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. Mol. Phylogenet. Evol. 9: 572–584.

Springer, M.S., Amrine, H.M., Burk, A., Stanhope, M.J. 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., DeBry, R.W., Douady, C., Amrine, H.M., Madsen, O., De Jong, W., Stanhope, M. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol. Biol. Evol. 18: 132–143.

Stanhope, M.J., Waddell, V.G., Madsen, O., De Jong, W., Hedges, S.B., Cleven, G.C., Kao, D., Springer, M.S. 1998. 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.

Strimmer, K., Von Haeseler, A. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13: 964–969.

Thompson, J.D., Higgins, D.G., Gibson, T.J. 1994. Improved sensitivity of profile searches through the use of sequence weights and gap exision. Comput. Appl. Biosci. 10: 19–29.

Waddell, P.J., Cao, Y., Hauf, J., Hasegawa, M. 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.

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