

Four New Mitochondrial Genomes and the Increased Stability of Evolutionary Trees of Mammals from Improved Taxon Sampling

Yu-Hsin Lin, Patricia A. McLenachan, Alicia R. Gore, Matthew J. Phillips, Rissa Ota, Michael D. Hendy, and David Penny

Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, New Zealand

We have sequenced four new mitochondrial genomes to improve the stability of the tree for placental mammals; they are two insectivores (a gymnure, *Echinosorex gymmurus* and Formosan shrew *Soriculus fumidus*); a Formosan lesser horseshoe bat (*Rhinolophus monoceros*); and the New Zealand fur seal (*Arctocephalus forsteri*). A revision to the hedgehog sequence (*Erinaceus europaeus*) is also reported. All five are from the Laurasiatheria grouping of eutherian mammals. On this new data set there is a strong tendency for the hedgehog and its relative, the gymnure, to join with the other Laurasiatherian insectivores (mole and shrews). To quantify the stability of trees from this data we define, based on nuclear sequences, a major four-way split in Laurasiatherians. This ([Xenarthra, Afrotheria], [Laurasiatheria, Supraprimates]) split is also found from mitochondrial genomes using either protein-coding or RNA (rRNA and tRNA) data sets. The high similarity of the mitochondrial and nuclear-derived trees allows a quantitative estimate of the stability of trees from independent data sets, as detected from a triplet Markov analysis. There are significant changes in the mutational processes within placental mammals that are ignored by current tree programs. On the basis of our quantitative results, we expect the evolutionary tree for mammals to be resolved quickly, and this will allow other problems to be solved.

Introduction

The evolutionary tree of mammals is rapidly being resolved with important agreement between nuclear and mitochondrial data sets (for example see Mouchaty et al. 2000b; Cao et al. 2000; Madsen et al. 2001; Murphy et al. 2001a; Phillips et al. 2001; Waddell, Kishino, and Ota 2001). Restricting ourselves, for the moment, to eutherian (placental) mammals, recent work shows them grouping strongly into four major clades (Waddell, Kishino, and Ota 2001; Lin, Waddell, and Penny 2002) (Until the root of the eutherian tree is unambiguous one group could be paraphyletic.) The groups are:

Xenarthrans (or Edentata—armadillos, sloths, and anteaters);
Afrotheria (including elephants, hyraxes, sea cows, tenrecs, armadillo, golden moles, and elephant shrews);
Supraprimates (primates, tree shrews, flying lemurs, rodents, and lagomorphs; see Lin, Waddell, and Penny 2002); and
Laurasiatheria consisting of the artiodactyls (including whales), carnivores (including pinnipeds), pangolins, perissodactyls, bats, and the Laurasiatherian insectivores (shrew, mole, and hedgehog). Hedgehog is usually included in this group on nuclear data but not on mitochondrial.

This Laurasian group (without the bats and the mole and shrew [lipotyphlan insectivores]) was first strongly supported by mitochondrial genomes (see, Xu and Arnason 1996). Pumo et al. (1998) reported that bats, based on a single complete mitochondrial genome, joined the mammalian tree just outside this group. Later

the mole, also based on a complete mitochondrial genome, was shown to occur in a similar position (Mouchaty et al. 2000a). This combined group of Ferungulates, bats, and core insectivores was named Laurasiatheria in Waddell, Okada, and Hasegawa (1999).

Our long-term goal is to get good estimates of the timing of divergence of the main eutherian lineages, particularly to estimate how many mammal lineages survived from the Cretaceous to the Tertiary (Hedges et al. 1996; Cooper and Penny 1997; Penny et al. 1999; Eizirik, Murphy, and O'Brien 2001). However, there are many potential sources of error in getting good estimates for early times of divergence (see Waddell and Penny 1996). A reliable evolutionary tree will remove a significant source of error, and it is therefore especially important to estimate the accuracy of evolutionary trees. The questions considered here are:

the position of hedgehog in relation to the mole and shrew,
whether megabats are derived from microbats; and
the relationship between the bats and Laurasiatherian insectivores (Lipotyphla).

The position of the hedgehog among eutherians has been problematic. Its mitochondrial genome was one of the earliest reported (Krettek, Gullberg, and Arnason 1995), and in most analyses the hedgehog appears as the first divergence in eutherians (for example, Krettek, Gullberg, and Arnason 1995; Penny et al. 1999). However, it was recognized early that the hedgehog mitochondrial genome had an anomalous nucleotide composition, including a high A + T content. Even when this is partly compensated for with the LogDet (paralinear) transformation (Lockhart et al. 1994), the hedgehog still appears as the outgroup to the remaining eutherian mammals. Rich et al. (1997), in reporting an early mammalian fossil, considered hedgehog to be an early eutherian, but this conclusion was partly based on the results from mitochondrial data. Lou, Cifelli, and Kielan-

Key words: bats, insectivores, mammal evolution, mitochondrial genomes, taxon sampling, tree comparisons, triple Markov analysis.

Address for correspondence and reprints: David Penny, Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand. E-mail: d.penny@massey.ac.nz.

Mol. Biol. Evol. 19(12):2060–2070. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Jaworowska (2001) reanalyzed the fossil data with new finds and did not come to the same conclusion. Analyses of nuclear sequences however, usually place the hedgehog closer to its supposed morphological relatives—within the Laurasiatherian insectivores (Lipotyphla), including shrew and mole (for example, Madsen et al. 2001; Murphy et al. 2001a—but see Springer et al. 1997). In particular, hedgehog was closer to shrew than to mole. (Shoshani and McKenna [1998] and Douady et al. [2002] summarize work on placental insectivores.) Given its unexpected positioning with mitochondrial data and its anomalous nucleotide composition, several authors have omitted the hedgehog from their analyses (for example, Mouchaty et al. 2000b; Reyes et al. 2000).

In the present work we reconsider the problem of hedgehog and lipotyphlan insectivores by sequencing two additional mitochondrial genomes (a gymnure and a shrew), and by resequencing some problematic portions of the hedgehog mitochondrial genome. The gymnure (moon rat, or hairy hedgehog) is in the same family (Erinaceidae) as the hedgehog but is in a different subfamily (Hylomyiinae rather than Erinaceinae, McKenna and Bell 1997). If there is a long-branch attraction problem (Hendy and Penny 1989) in relation to hedgehog, then a combination of mitochondrial genomes from a gymnure (in a related subfamily) and a shrew (in a related family) has an improved chance of resolving the hedgehog position.

Part of the hedgehog mitochondrial DNA was resequenced because we noted that the protein-coding alignment showed a number of regions where 5–12 consecutive amino acids, that were normally conserved, had substitutions. From inspection of the DNA sequence we suspected that these anomalous regions were the result of either three successive single-base insertions or an insertion with a later complementary deletion (thus re-establishing the reading frame). The reading frame was interrupted for 15–36 nucleotides. M. D. Sorenson (personal communication) has also noted the same phenomenon. Such an interruption of reading frame obviously introduces errors into the data set. However, these are expected to be random in occurrence and, as evaluated by simulation studies (Charleston 1994, pp. 115–131), random errors have little effect on the accuracy of recovering the correct tree, especially as compared with systematic errors. Nevertheless, it is highly desirable to eliminate all sources of errors, and therefore parts of the hedgehog were resequenced.

In addition to the hedgehog–lipotyphlan insectivore question, there are uncertainties over the relationship both between microbats and megabats, as well as between bats and the Laurasian insectivores. The monophyly of bats seems well established (see Lin and Penny 2001), but it has recently been suggested (Hutcheon, Kirsch, and Pettigrew 1998; Teeling et al. 2000) that megabats are derived from within microbats. In particular, megabats appeared closer to rhinolophid (horseshoe and false vampire) microbats. In this model, megabats are strictly monophyletic and microbats paraphyletic. Mitochondrial genomes have not previously been available for Rhinolophid microbats (except 12S–16S

rRNA). The two microbat genomes previously available had been the Jamaican fruit bat (in the family Phyllostomidae, Pumo et al. 1998) and New Zealand long-tailed bat (in the family Vespertilionidae, Lin and Penny 2001). Recently, three other Laurasiatheria mt-genomes have become available in Nikaido et al. (2001). These are pipistrelle and rhinolophid bats and a shrew. Our four new sequences are analyzed with these three new mt-genomes in a Laurasiatherian data set, but they are not included in the full data set of mammals. Finally, including taxa with good fossil records is important for estimates of the timing of divergence of the main eutherian lineages. Fur seal will help in future as an important calibration point for dating now that bear mt-genomes are available (Delisle and Strobeck 2002).

With improved taxon sampling, and with the exception of the hedgehog that is being studied here, there is good basic agreement between trees from mitochondrial data sets and from nuclear data sets. However, to formalize this we require quantitative measures of the similarity of trees (Steel and Penny 1993). A standard criticism, for example Goldman, Anderson, and Rodrigo (2000), is that tests of significance for trees (such as the Kishino-Hasegawa test) are designed for evaluating pre-determined hypotheses (trees). In contrast, virtually all phylogenetic studies do the opposite; they infer trees from the data as if no prior knowledge (hypotheses) was available and then start testing the resulting hypotheses. Ota et al. (2000), Goldman, Anderson, and Rodrigo (2000), Waddell, Kishino, and Ota (2001), and Strimmer and Rambaut (2002) discuss some of the problems. Here we compare trees from different data sets because this allows other hypotheses to be tested for the overall similarity of trees derived from independent data sets (Penny, Foulds, and Hendy 1982; Steel and Penny 1993).

Materials and Methods

Samples of gymnure (*Echinosorex gymnurus*) were provided by Adura Mohd Adnan, Malaysia. Cheng Hsi-Chi, Taiwan, supplied the Formosan shrew (*Soriculus fumidus*) and a Formosan lesser horseshoe bat (*Rhinolophus monoceros*). The New Zealand fur seal sample (*Arctocephalus forsteri*) was supplied by Pdraig Duignan of the Massey Veterinary School, sample SS9771AF. The hedgehog was from the New Zealand population introduced from England (Wodzicki 1950) and is subspecies *Erinaceus europaeus europaeus*.

DNA was extracted from muscle or liver using High Pure[™] PCR Template Purification Preparation Kit (Roche). With all samples, mitochondrial DNA was amplified in fragments longer than 5 kb (to avoid amplifying nuclear copies) using the Expand[™] Long template PCR kit (Roche). Long PCR DNA fragments were sequenced directly and also used as template for a second short range PCR of 1–2 kb. Sequencing reactions were done according to standard protocols and run on a 377 ABI Applied Biosystems DNA sequencer. Because we are sequencing several complete mt-DNA genomes, we designed primers from conserved regions of the mt-DNA genomes of mammals and birds, allowing 0–5 de-

generate sites to maximize their usefulness for other species. We used the Fasta search in the GCG program (Wisconsin Package, version 10.0) to search our primer database for appropriate targets for primer walking. When none were available, new primers were designed using Oligo[®]4.03 (National Biosciences, Inc.). Sequences were checked and assembled using Sequencing Analysis and Sequence Navigator programs (ABI) and Sequencher.

Three sets of sequences were used for analysis: 29 Laurasiatherians, 42 eutherians, and 47 mammals. Each larger data set included all taxa from the smaller data sets. Complete mammalian mt-DNA sequences were obtained from GenBank for the following Laurasiatheria taxa: Jamaican fruit bat *Artibeus jamaicensis* [NC_002009]; Ryukyu flying fox *Pteropus dasymallus* [NC_002612]; mole *Talpa europaea* [NC_002391]; hedgehog *Erinaceus europaeus* [NC_002080]; dog *Canis familiaris* [NC_002008]; cat *Felis catus* [NC_001700]; harbor seal *Phoca vitulina* [NC_001325]; gray seal *Halichoerus grypus* [NC_001602]; horse *Equus caballus* [NC_001640]; donkey *Equus asinus* [NC_001788]; white rhinoceros *Ceratotherium simum* [NC_001808]; Indian rhinoceros *Rhinoceros unicornis* [NC_001779]; cow *Bos taurus* [NC_001567]; sheep *Ovis aries* [NC_001941]; fin whale *Balaenoptera physalus* [NC_001321]; blue whale *Balaenoptera musculus* [NC_001601]; sperm whale *Physeter catodon* [NC_002503]; hippopotamus *Hippopotamus amphibius* [NC_000889]; pig *Sus scrofa* [NC_000845]; and alpaca *Lama pacos* [NC_002504]. Three recently published mt-genomes are the Japanese pipistrelle bat (*Pipistrellus abramus* [AB061528]), rhinolophid (*Rhinolophus pumilus* [AB061526]), and a long-clawed shrew (*Sorex unguiculatus* [AB061527]). Two new sequences were available from our laboratory for bats, an Australian flying fox *Pteropus scapulatus* [NC_002619] and the NZ long tailed bat *Chalinolobus tuberculatus* [NC_002626] (Lin and Penny 2001). The four new genomes reported give a data set with 29 mitochondrial genomes from distinct Laurasiatherian species (assuming in the interim that hedgehog fits within this group). For this Laurasian set we predicted an expected tree (unrooted) from the most recently published nuclear data (Eizirik, Murphy, and O'Brien 2001; Madsen et al. 2001; Murphy et al. 2001a). Where nuclear sequences were not available (such as for gymnure) we used the accepted classification based on morphological characters. We predicted this tree from nuclear data would be extremely similar to the optimal tree from the slowest evolving character states of the mitochondrial data.

To help identify the root of the Laurasian grouping, an expanded data set was made with a wide range of 16 other eutherians. These were mouse *Mus musculus* [NC_001569]; red squirrel *Sciurus vulgaris* [NC_002369]; guinea pig *Cavia porcellus* [NC_000884]; fat dormouse *Myoxus glis* [NC_001892]; cane rat *Thryonomys swinderianus* [NC_002658]; rabbit *Oryctolagus cuniculus* [NC_001913]; human *Homo sapiens* [NC_001807]; baboon *Papio hamadryas* [NC_001992]; white-fronted capuchin *Cebus albifrons* [NC_002763]; slow loris *Nycti-*

cebus coucang [NC_002765]; aardvark *Orycteropus afer* [NC_002078]; elephant *Loxodonta africana* [NC_000934]; tenrec *Echinops telfairi* [NC_002631]; and armadillo *Dasyurus novemcinctus* [NC_001821]. In addition, two sequences were available within the laboratory, a pika *Ochotona collaris* [AF348080] and a vole *Volemys kikuchii* [AF348082] (Lin, Waddell, and Penny 2002), giving a total of 42 mitochondrial genomes of eutherian mammals (using 26 Laurasians and 16 others). Additional sequences, such as other apes, including chimpanzee, gorilla, and orangutan, were not used because they are all close to the human sequence and do not help resolve the deeper eutherian divergences. The tree shrew is analyzed in Lin, Waddell, and Penny (2002) and is variable within Supraprimates but does not help with Laurasiatherians.

Four mitochondrial genomes are available for marsupials including the previously published sequences for opossum *Didelphis virginiana* [NC_001610] and wallaroo *Macropus robustus* [NC_001794], together with two from within our laboratory (Phillips et al. 2001), a bandicoot *Isoodon macrourus* [NC_002746] and a brush-tailed possum *Trichosurus vulpecula* [AF357238]. The platypus *Ornithorhynchus anatinus* [NC_00089] was also used. These five taxa were combined with the sequences in the eutherian data set to give the "mammalian" data set.

To increase our ability to compare results quantitatively, we prepared four subsets for each of the Laurasiatherian, eutherian, and mammalian data sets. The subsets contained the following

1. RNA sequences (rRNAs and tRNAs),
2. First and second nucleotides from 12 protein genes coded on the same DNA strand,
3. combined RNA-protein data set (1 and 2), and
4. protein data as amino acids.

Thus we could compare results for 12 sets of data, four subsets for each of the three data sets (Laurasiatheria, eutherian, and mammalian). Sequences were aligned manually in Se-Al version 1.0a1 (<http://evolve.zps.ox.ac.uk/Se-Al/Se-Al.html>). The rRNA sequences are aligned with reference to the secondary structure (<http://www.rna.icmb.utexas.edu/RNA/>) to maximize homologous positions. Data sets are available from (<http://awcmee.massey.ac.nz/downloads.htm>).

PAUP* 4d65 (Swofford 1998) was used for all data sets. MOLPHY (Adachi and Hasegawa 1996) was used for a maximum likelihood analysis of amino acids sequences. A triplet Markov analysis (analyzing three sequences simultaneously, rather than pairs of sequences) was undertaken by the procedure of Lake (1997), which is similar to that of Chang (1996). This estimates the Markov transition matrices from the root to each of the three lineages; we used an updated version of the Bootstrappers gambit program available from <http://www.mcdb.ucla.edu/Research/Lake/Research/Programs/>. The smaller Laurasian data set was analyzed first, to obtain the unrooted tree for Laurasiatherians. Then the eutherian data set was analyzed both to identify the root of the Laurasian tree, and check whether hedgehog (and gymnure) stayed within the Laurasian group. Finally, the

Table 1
Estimated Markov Transition Matrices from a Root to Each of Three Insectivores

	ROOT TO GYMNURE				ROOT TO SHREW				ROOT TO MOLE			
	T	C	A	G	T	C	A	G	T	C	A	G
T ...	0.942	0.146	0.063	0.019	0.967	0.059	0.019	0.009	0.969	0.036	0.021	0.010
C ...	0.028	0.796	0.015	0.006	0.020	0.928	0.015	0.004	0.021	0.949	0.015	0.003
A ...	0.026	0.053	0.906	0.097	0.010	0.012	0.955	0.034	0.008	0.013	0.947	0.014
G ...	0.003	0.006	0.015	0.878	0.002	0.002	0.011	0.953	0.002	0.002	0.016	0.973

NOTE.—The matrix is read from top right to left.

full mammalian data set was studied to check the rooting of the eutherian tree and to detect whether adding the outgroup (marsupials plus platypus) led to any rearrangements within the eutherians (such as those that can arise from the long edges attract phenomenon, Hendy and Penny 1989).

Results

Our new mitochondrial genomes are available from GenBank, numbers AF348079 (gymnure), AF348081 (shrew), AF406806 (rhinolophid bat), and AF513820 (fur seal). The sequences have the standard gene order of mammals, and are 17,088, 17,488, and 16,851 nucleotides long for the gymnure, shrew, and horseshoe bat, respectively. The fur seal control region is not complete but does an 8-bp gatacac tandem repeat, almost identical to that of the harbor seal. Gymnure has a tandem repeat cagcta, shrew has cacgtata, and hedgehog catacg. The similarity of the tandem repeat does not guarantee homology, for example, New Zealand long-tailed bat and the little red flying fox also have catacg. The gymnure does show a high thymine-cytosine ratio, similar to the hedgehog (see Phillips et al. 2001). The T-C ratio averages 1.05 for 5365 variable sites from all protein-coding and RNA genes and for 50 mammalian mitochondrial genomes. It is 1.43 for hedgehog and gymnure (it is also relatively high for *Didelphis* and the bandicoot—1.27 and 1.31, respectively). Thus this unusual feature of the hedgehog mt-genome occurs in both subfamilies of the Erinaceidae. As mentioned above, the gymnure sequence has this same unusual nucleotide composition as the hedgehog, though it has not evolved

quite as fast as the hedgehog. Apart from this high thymine-cytosine ratio, the genomes do not show any unusual features. The new hedgehog sequences have the following GenBank numbers, AF513817-AF513819. They are for the complete NADH2 gene and are for parts of the COIII and NADH4 genes. They confirm that the original sequence had some small insertions and deletions, the relevant sections are shown as Supplementary Information. M. D. Sorenson (personal communication) has also resequenced this portion of the hedgehog and obtained the same result.

Current methods for inferring evolutionary trees generally assume the same process occurs across the tree, that is, the process is stationary. We have already used a triplet Markov analysis to present evidence that the murid rodents have a different mutational process to other rodents, and to most other placental mammals (Lin, Waddell, and Penny 2002). This compares three sequences at a time using a tensor (three-dimensional array) which records the frequencies of nucleotide triplets across the three sequences. There is sufficient information in the tensor to recover the three 4×4 Markov transition matrices from the root to each of the three taxa. Results are shown for the protein- and RNA-coding section of the complete mt-genomes of gymnure, mole, and shrew in table 1. The important point in the present context is that it demonstrates that there is a change in the mutational process on the gymnure and hedgehog lineage. Such a “change in mutational process” violates the assumptions of most methods of analysis. Consequently, extra care is required in interpreting any unexpected results.

Laurasiatheria Data Set

Before considering the tree for just the Laurasiatheria we will give our predictions (based on the trees from nuclear data of Madsen et al. 2001; Murphy et al. 2001a; Waddell, Kishino, and Ota 2001) for the four deepest splits in Laurasiatheria. These are the Laurasiatherian insectivores (Lipotyphla), bats, whales plus artiodactyls, and Carnivora plus perissodactyls (this latter being the weakest prediction, artiodactyls and perissodactyls sometimes come together). The bats and Lipotyphla are expected to be adjacent on the unrooted tree (see fig. 1). In the Laurasiatherian data there are 5, 7, 8, and 9 taxa in these groups. The probability of selecting a random binary phylogenetic tree with the taxa partitioned into these four groups and connected as in figure 1 is 1.89×10^{-18} . In general, if the n taxa are

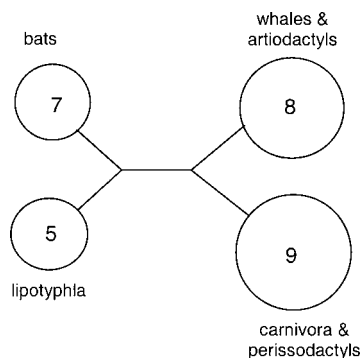


FIG. 1.—Predicted relationship between four groupings of Laurasiatherians. In the data set used there are seven bat mitochondrial genomes, five lipotyphla, nine carnivores and perissodactyls, and eight whales plus ungulates.

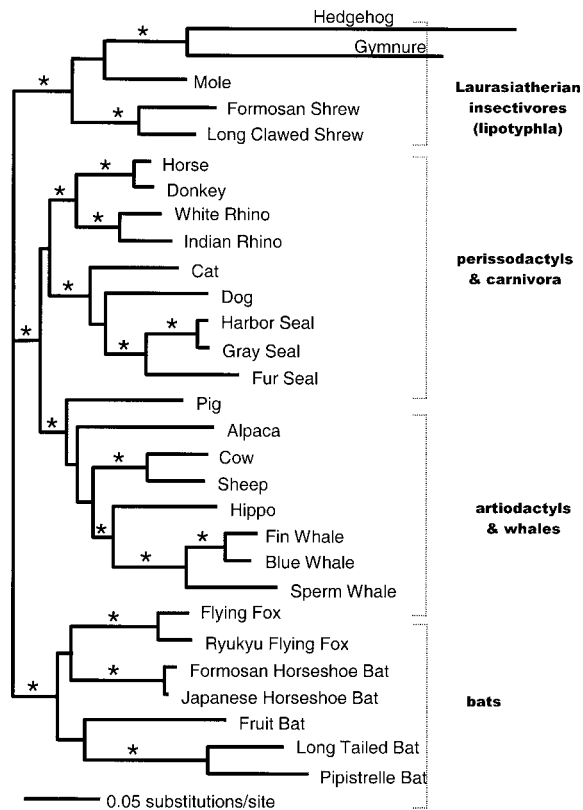


FIG. 2.—The unrooted maximum likelihood tree for 29 Laurasiatherian taxa; combined RNA and 1 + 2 sites of protein-coding genes and using HKY85 model with invariant sites estimated. The relationships predicted in figure 1 are found in this tree. Bootstrap values of ≥ 99 are marked with an asterisk.

partitioned into subsets with w , x , y , and z taxa, then there are $3b(w + 1) \cdot b(x + 1) \cdot b(y + 1) \cdot b(z + 1)$ ways of connecting them to form a binary tree (see *Appendix*), where $b(n)$ is the number of unrooted binary trees connecting n taxa. Hence the probability of a specific connecting arrangement is $b(w + 1) \cdot b(x + 1) \cdot b(y + 1) \cdot b(z + 1) / b(n)$. This approach uses independent data sets (in this case mitochondrial and nuclear) to quantitatively estimate the similarity of the predicted results.

Figure 2 shows the maximum likelihood tree for the combined RNA and codons 1 + 2 Laurasiatherian data set, and it has the basic arrangement (fig. 1) as found in nuclear data sets. Even if we reduce the data set by considering only a single megabat, equid, rhinoceros, seal, whale, and one from sheep and cow, the probability of randomly selecting the tree with four, four, five, and five members in each subtree is only $\approx 1.29 \times 10^{-11}$. These results are the basis of our claim that there is a fundamental similarity between “most” results on mitochondrial and nuclear data sets and that therefore the mammalian tree is rapidly being resolved. The result should not be overinterpreted as implying the relationship is “correct”—there could be other trees almost as good. The importance is in demonstrating that there is strong, and congruent, information in both mitochondrial and nuclear data. We will see later that there are difficulties with specific taxa that current methods are not designed to handle correctly, but finding high basic

congruence is excellent, and shows the power of molecular approaches to studying evolution.

The tree in figure 2 is from MOLPHY, the bootstrap values are from ML distances and neighbor joining. The 18 bootstrap values of 99% or greater in figure 2 are marked by an asterisk, and the other two are 97% and 98%. The average is 95% and the lowest 56% (within the artiodactyls, between alpaca and cow-sheep). We also carried out a nearest-neighbor analysis on the bootstrap results (see Cooper and Penny 1997) to measure local stability on the tree. This sums all the values within a single interchange around each internal branch of the tree. The average value is now 99% with the lowest value of 92%. Thus, 99% of bootstrapped trees have no more than simple rearrangements around internal edges of the tree. Most results therefore are within just a single interchange on the optimal tree; the tree is locally stable in our terminology (Cooper and Penny 1997).

Of the four new mitochondrial genomes, the position of the fur seal is quite straightforward and is considered first. The seals (Pinnipedia) have three extant families: Odobenidae (walrus), Otariidae (fur seals), and Phocidae (including gray seal and harbor seal). In all our results, the position of fur seal with respect to the gray seal and harbor seal is stable (figs. 2–5) and thus supports this traditional taxonomy of the group. The relationship between these three families is unknown (Lento et al. 1995) and a complete walrus mt-genome is now required. It is also desirable to have other members of the dog group of Carnivora (bears, pandas, and ferrets–otters are considered more closely related to seals than dogs) and this will then give an additional calibration point on the eutherian tree (Berta, Ray, and Wyss 1989). Similarly, another member of the cat group, such as a hyena, mongoose, or viverrid is desirable because the bootstrap value for the cats versus dogs is one of the lowest on the Laurasian data set, about 75% with MOLPHY. Our horseshoe (Rhinolophid) microbat is very similar to that of Nikaido et al. 2001, however, our data set now has seven bats (rather than four), and this gives more stability within the bats. In the studies of Hutcheon, Kirsch, and Pettigrew (1998) and Teeling et al. (2000) the rhinolophoid bats are the sister group of megabats—resulting in megabats being monophyletic and microbats being paraphyletic. Our results support this hypothesis (see table 2 for Laurasian data set); results for the 11 other data sets are summarized in table 2. Although the maximum likelihood on the combined data, and on amino acid sequences, give megabats as derived from microbats, it is still desirable to have some additional bat sequences—a megadermatid bat (a ghost bat or a false vampire bat) and a distant megabat for example, to strengthen these conclusions.

In all data sets and with all methods of analysis the gymnure and hedgehog come together, and in the Laurasian data set they are always with the other insectivores (the two shrews and mole). Similarly, the two shrews (from different genera) are united in all data sets. There is some variation in relative positions of the shrews, mole, and (hedgehog, gymnure), but the mutational process in the hedgehog–gymnure lineage is so different

Table 2
Comparative Results by Data Set and Method of Analysis

Subset→		1+2			RNA			(1+2)+RNA			amino acids		
		ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ
Laur	Ins-Mon*	✓b	✓b	✓b	✓b	✓c	✓b	✓b	✓c	✓b	✓b	✓a	✓b
	R+Mbat	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
Euth	Ins-Mon	×A	×	×	✓a	✓b	✓b	✓a	×A	×C	×A	×A	×A
	R+Mbat	✓	✓	✓	✓	×	✓	✓	×	✓	✓	✓	✓

NOTE.—Data set: Laur = 29 Laurasians (fig. 2), Euth = 42 Eutherians (placentals) (fig. 3), Mamm = 47 Mammals (figs. 4 and 5). Ins-Mon: Insectivora (Lipotyphla) monophyly. R + Mbat Rhinolophorus bat and megabat are sister taxa. ML: ME with ML (HKY85) distances. MP: Maximum Parsimony. NJ: Neighbor joining, LogDet + Invariant sites model. a: ((hedgehog, gymnure), shrew), mole). b: ((hedgehog, gymnure), (shrew, mole)). c: ((hedgehog, gymnure), mole), shrew). A: hedgehog–gymnure join mouse–vole branch. B: hedgehog–gymnure are the deepest branch (after marsupials). C: hedgehog–gymnure are only one step from shrew–mole.

* Lipotyphla monophyly is dependent on the rooting of the tree. If the root is, for example, between hedgehog–gymnure and mole–shrew, then lipotyphla becomes paraphyletic.

that little confidence can be placed in their precise relationship until better analytical methods are possible. At this point we add other eutherian mammals into the data set to test whether this insectivore group stays together.

Eutherian Data Set

In trees with the Laurasiatheria data set, our results are consistent with monophyly of Lipotyphla (shrews,

mole, and hedgehogs). But strict monophyly is dependent on the rooting of the Laurasiatheria, especially in relation to the position of hedgehog and gymnure. Similarly, when only a single sequence was available for each of the bats and the core insectivores (the mole), these tended to form sister groups (for example, Mouchaty et al. 2000b). However, as additional sequences of bats became available there was a tendency (for example, Lin and Penny 2001) for the insectivores to diverge first, then the bats, and finally the Ferungulates (which now includes whales). Thus it is necessary to estimate the position of the root of the Laurasiatherian tree—as a first step the data set with 42 eutherians is used. The maximum likelihood tree on the combined RNA and protein-coding genes uses 16 other placentals to root the Laurasiatherian tree (fig. 3). There are no surprises among the outgroup taxa. Armadillo and the Afrotherians (elephant, aardvark, and tenrec) are united, as are Supraprimates (primates, rodents, and lagomorphs). The basic four-way division of eutherians ([Xenarthra, Afrotheria], [Supraprimates, Laurasiatherians]) is found here as well (see Lin, Waddell, and Penny 2002). This fundamental division appears very robust, especially in large data sets.

On either the RNA or the combined protein + RNA data set, the hedgehog and gymnure can still combine with the mole and the two shrews giving the Lipotyphla (fig. 3 and table 2). This is the first time that hedgehog has been grouped with mole based on mt-genomes. Traditionally, the Lipotyphla includes two suborders, Soricomorpha (including mole in the family Talpidae and shrew in the family Soricidae) and Erinaceomorpha (including hedgehog and gymnure, both in the family Erinaceidae, see Butler 1988 and MacPhee and Novacek 1993). However, recent publications, mainly from nuclear genes, placed moles as sister taxa to both shrew and hedgehog (Eizirik, Murphy, and O'Brien 2001; Murphy et al. 2001a). Our results were ambiguous in their relationship; Soricomorpha can be monophyletic or paraphyletic in different data sets and analysis methods (see table 2). As mentioned earlier, because the hedgehog and gymnure have such an anomalous nucleotide composition any detailed results with them should be treated with caution until better methods of analysis are available (that is, methods that incorpo-

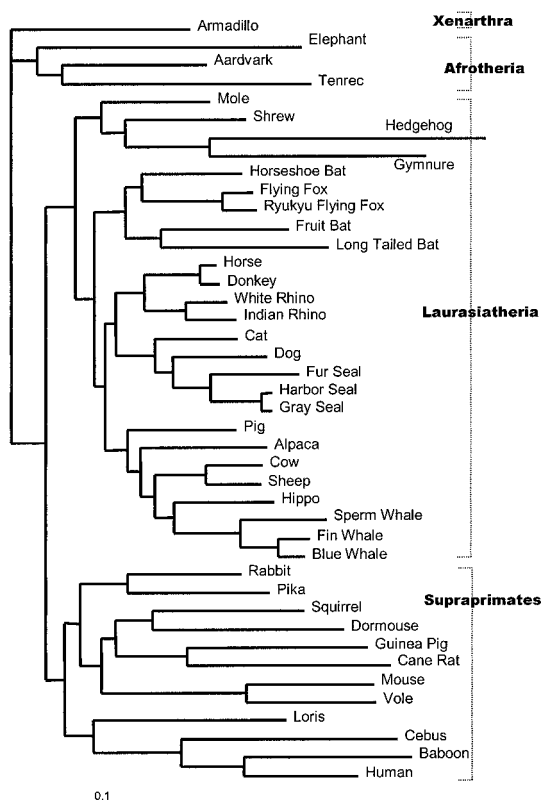


FIG. 3.—The unrooted maximum likelihood tree for 42 eutherians; combined RNA and sites 1 + 2 of protein-coding genes. The Laurasiatherians are rooted with the core lipotyphlan insectivores on one side and bats plus ferungulates on the other. The hedgehog and gymnure form the Lipotyphla with mole and shrew. The horseshoe (Rhinolophid) microbat is a sister group to the megabats (flying foxes). This ingroup (Laurasiatherian) tree should be stable when the marsupial-platyputus outgroup is added (see figs. 4 and 5).

rate the unusual features of these sequences). With the protein subset of the eutherian data set the murid rodents (such as mouse and vole) tend to join with the hedgehog and gymnure. The murid rodents also have an anomalous mutation process (see discussion in Lin, Waddell, and Penny 2002). Previous work had focused on the guinea pig as being unusual, but with more rodent sequences now available, and with extensive information on DNA repair in murid rodents, it is clear that it is the murids that are showing atypical behavior (Lin, Waddell, and Penny 2002).

On the combined RNA and protein-coding data set, Lipotyphla is deepest in the Laurasian group. They are followed by Chiroptera (table 2) and with high bootstrap support (91% in amino acid ML). Thus the single long-branch of a mole and a bat in Mouchaty et al. (2000b) may be a consequence of long-branch attraction. However, recent studies of nuclear genes show that the relative positions of insectivores and bats still varies slightly within Laurasiatheria (Eizirik, Murphy, and O'Brien 2001; Madsen et al. 2001; Murphy et al. 2001a; Waddell, Kishino, and Ota 2001). Overall we find that as additional bat and insectivore sequences are added the insectivores tend to diverge earliest, rather than forming a sister group with bats. Given that a major feature of our results is that additional taxa are reducing the long-branch attraction problem (Hendy and Penny 1989), the conclusion that Lipotyphla diverge first is our best estimate at present. Future work needs to combine all the data sets, mitochondrial and nuclear to check this result.

Mammalian Data Set

The root of the eutherian tree is still not adequately resolved (see Murphy et al. 2001b; Waddell, Kishino, and Ota 2001). Adding an outgroup of four marsupials plus platypus should help narrow down the position of the root. In general, adding an outgroup to the ingroup tree is the most difficult part of a study to get correct because the outgroup (by definition) is furthest away, and any slight changes in the evolutionary process will be magnified relative to the ingroup (Holland et al., in preparation). This effect is exaggerated if only a single outgroup sequence is used (here we use five). However, if the model of evolution used is correct, then adding an outgroup should not result in any changes within the ingroup.

However, we find a major rearrangement in the eutherian tree when the five outgroup taxa are added into the data set (fig. 4). The gymnure–hedgehog pair moves seven steps on the tree in figure 3 becoming adjacent to murid rodents, and simultaneously (because the tree is now rooted) becomes the deepest branch in the eutherians. *The trees in figures 3 and 4 cannot both be correct!* The simplest explanation is that there is an error in the assumptions about the mechanism of evolution. The unexpected relationships in figure 4 are suspicious because of the abnormal base frequency of gymnure and hedgehog, differences in DNA repair in murid rodents (Lin, Waddell, and Penny 2002), and the unusual nucle-

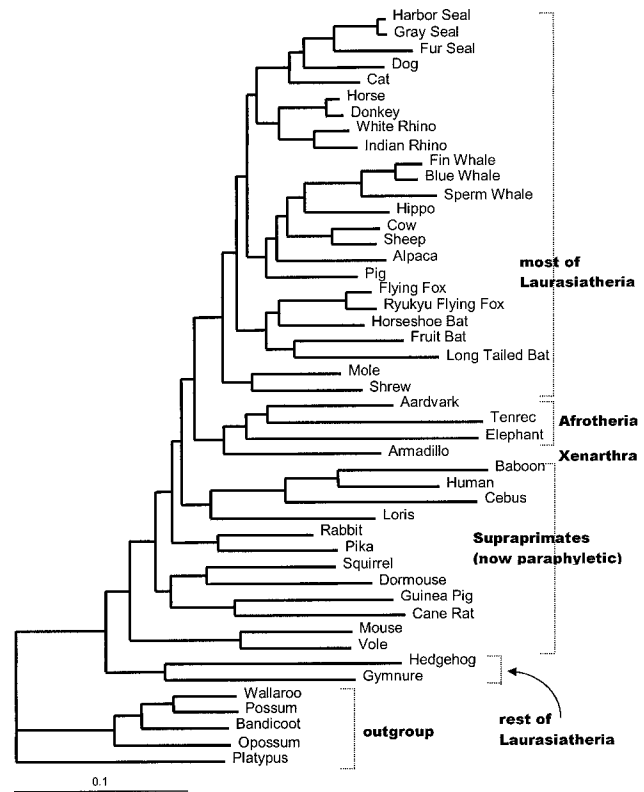


FIG. 4.—The mammalian tree with eutherians rooted by the four marsupials plus platypus. Hedgehog plus gymnure move considerably within eutherians, separating from mole and shrew to join to the mouse–vole group. Thus adding the outgroup (marsupials) causes major rearrangements within the ingroup tree, contradicting the expectations from the model. Hedgehog and gymnure shift seven steps along the tree in figure 3 to become adjacent to the vole–mouse. This rearrangement of the ingroup is characteristic of the classical long-branch attraction reported in Hendy and Penny (1989) where the outgroup in the five-taxon equal-rate example led to a rearrangement within the ingroup tree. The effect of constraining the hedgehog–gymnure is shown in figure 5.

otide compositions in some marsupials (Phillips et al. 2001).

When using marsupials as the outgroup to placentals, Lipotyphla became polyphyletic, rodents paraphyletic, and hedgehog–gymnure become the first branches to the rest of placentals (fig. 4). This is in contrast to morphological and recent molecular (nuclear genes) studies (Butler 1988; Eizirik, Murphy and O'Brien 2001; Madsen et al. 2001; Murphy et al. 2001a).

Figure 5 is obtained when we constrained the tree to retain lipotyphlan monophyly. The root of the placental tree now joins between the Xenarthran–Afrotheria and the Laurasiatherian–Supraprimate groups. The Laurasian part of the tree is now consistent with the earlier unrooted tree of figure 3, with the root sitting between Lipotyphla and the rest of the Laurasians. If the tree in figure 4 had been correct, then the root in figure 5 should still have come into either the murid rodents or the gymnure–hedgehog lineage. On the Kishino–Hasegawa test (KH-test) the ML tree without constraint (fig. 4) was not significantly better than the constrained tree in figure 5 ($P \cong 0.2–0.7$). In NJ using HKY85+I+ Γ model with

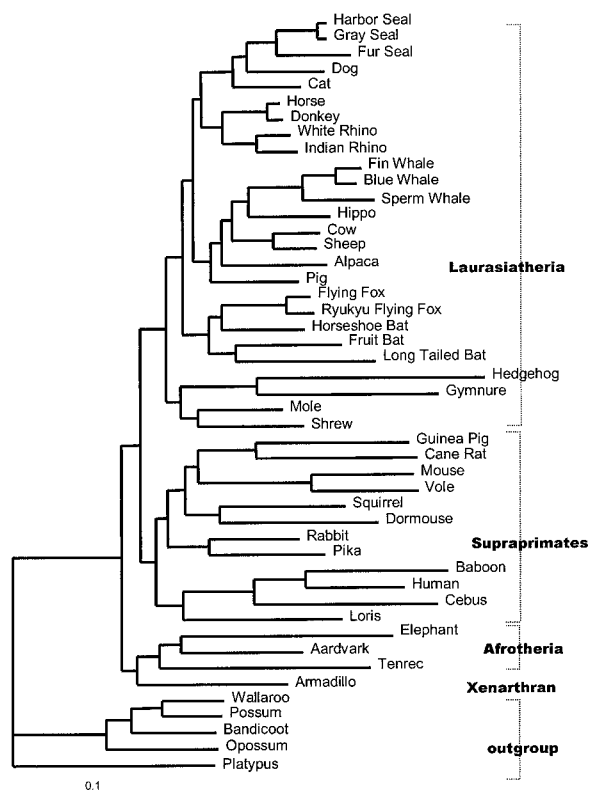


FIG. 5.—The tree with hedgehog–gymnure–mole–shrew constrained together. The eutherian root now separates the Xenarthran (armadillo) and the three Afrotherians from the Laurasiatherians and Supraprimates, and the ingroup tree is consistent with figure 3.

48% of invariant sites removed (estimated using ML) and α value in 0.2–1.0, the tree inferred is similar to figure 5, and showed Lipotyphla monophyly. Thus given all the information available, we consider the tree in figure 5 as our best estimate at present, though it is desirable to have additional mt-genomes. Perhaps a rodent basal to the murid rodents would be the easiest to obtain.

Discussion

With a larger number of taxa in the data set, there is good agreement between trees from independent data sets. Trees from nuclear and mitochondrial data are highly similar, as are trees from RNA and protein coding regions of mitochondria (see also Waddell, Kishino, and Ota 2001). There are specific problems for murid rodents (that may be solved now that a *Spalax* mt-genome has been sequenced, A. Reyes, personal communication), the hedgehog family, and for some genes in primates (Andrews and Easteal 2000). The three primates in Arnason et al. (2001) removed some earlier problems we had from the long-branch leading to primates. Recent progress has come from the experimental side by collecting additional data, rather than from the theoretical side of improving models to be more realistic and robust! Our conclusion (based on real data) contradicts that of Rosenberg and Kumar (2001) from simulation studies that incomplete taxon sampling is not a problem

for phylogeny reconstruction. Simulations generally use the same mechanism to generate data and to infer a tree, and use random models of speciation—neither of which holds with real data.

Several methods are used here to help understand the similarity of trees from different data sets. Predefining an expected tree (or subsets in the tree) on data previously available helps, and the probability of finding such subsets in new data can be calculated. Logically, which data set is used for the prediction, and which for testing, does not matter. Similarly, building a tree on the ingroup, then adding outgroups and checking for any changes to the ingroup, is essential in any analysis. The classical five-taxon case (Hendy and Penny 1989) has an ingroup in a zone of consistency—until the outgroup is added. We have recently reanalyzed this case (Holland, Huson, Penny, and Hendy, in preparation) and find many examples where adding the outgroup disrupts the ingroup. (In principle, adding outgroups could “fix” an error in the ingroup, but they are rare with simple examples.) Finally, the triplet Markov analysis (Chang 1996; Lake 1997) has potential for detecting deviations from the assumptions of the standard models. So, although the gains in understanding mammalian evolution come from the experimental side, there is still plenty of potential on the theoretical side.

Two aspects where our models of molecular evolution are inadequate are the mechanisms used by models and the genuine changes in mutational processes. Within eutherians two interesting examples are a change in mutational mechanism are among murid rodents and in hedgehogs and their relatives (including gymnures). With murids there is abundant evidence from cancer research that DNA repair is not as efficient in murids as in humans (see Lin, Waddell, and Penny 2002). Mutation includes both errors during DNA replication and mutations in nonreplicating cells that are not repaired (Huttley et al. 2000). The effects occur in both nuclear and mitochondrial genomes (not necessarily equally), and appear to have affected the rooting on nuclear data in both Springer et al. (1997) and Douady et al. (2002). There is considerable interest in changes in “rates” of evolution, with new methods to estimate times of divergence (for example, Kishino, Thorne, and Bruno 2001). There is less progress on testing that assumptions about the mutational process are indeed accurate. We distinguish between changes in the “rate” of evolution (all mutations increase or decrease by a similar proportion), and a change in “process.” Here there is differential acceleration (or deceleration) in mutation rates between pairs of nucleotides—thus leading to changes in nucleotide and dinucleotide frequencies (see for example, Karlin and Mrázek 1997).

There is a tendency to “blame the data” if we get an incorrect tree—the methods for inferring trees are assumed correct, the data is wrong. We take the opposite view—the data is correct and the methods of analysis are inadequate. It is the responsibility of theoretical biologists to develop robust methods that accurately reflect the processes in evolution. It is necessary to detect and then adjust for a change in process. The triplet Mar-

kov analysis (see results in table 1) is a useful step in detecting changes from analyzing the DNA sequences directly (before tree-building), though we have not yet estimated the bias on the values. Eventually we should be able to identify organisms for studying changes in DNA repair enzymes, thus helping understand the processes of molecular evolution better. We expect many signals in DNA, from history (phylogeny), multiple changes, changes in mutation processes, changes in functional constraint on proteins, positive selection, etc. (Penny et al. 1993). It is generally assumed that the signal from shared history (phylogeny) is the largest one, but there is no evolutionary reason to assume this. Organisms do not evolve “in order to” allow their history to be recovered, there are other processes that need to be dissected out and analyzed, mutational process is an important one.

To return to the data—updating the hedgehog sequence is useful, but by itself the new sequence does not lead to any changes in the tree. This is as expected because the errors lead to random, not systematic, errors. From similar experience with aligning it is likely that there are errors in other mitochondrial genomes that were sequenced early. In our experience Norway rat and *Xenopus* need some resequencing. The new *Rana* sequence [NC_002805] helps to some extent. It is not surprising that the first mitochondria sequenced had some errors simply because they had no close relatives for comparison. We still find during alignment that we detect an occasional error in our own results—and there are now far more genomes available for comparison. It was far more difficult for the first genomes sequenced.

With the Laurasiatheria, the composition of the main groups is relatively stable—though the position of carnivores with respect to perissodactyls requires more study. Two groups missing for mitochondrial data are the insectivore *Solenodon* and a pangolin. The former is placed in the lipotyphlan insectivores in Stanhope et al. (1998). Pangolins are close to carnivores on nuclear data (Madsen et al. 2001; Murphy et al. 2001a; Waddell, Kishino, and Ota 2001), and thus it is expected that pangolin will be with the group when a mitochondrial genome is available. Although we are emphasizing the high agreement of the trees between data sets, there is certainly local uncertainty in the eutherian tree. We define an evolutionary tree to be “locally stable” when any changes are limited to rearrangements around single internal edges (branches) (Cooper and Penny 1997). The relationship between bats and core insectivores is such an example—whether Lipotyphla are the first divergence or whether they form a sister group to bats. Both divergences are just a single interchange on the tree. With more sequences becoming available the pig appears one step deeper on the tree than alpaca, but even if this changes it is still a local rearrangement in our terminology. To conclude, obtaining more data for mammalian groups has given the major gain in understanding eutherian evolution. Our results illustrate the need for a similar gain in understanding the models of evolution. But it has been improvements in the data set rather than more elaborate computer programs, which have led to

our gain in understanding. It is certainly expected that as we go deeper into the tree of life that improvement in models will be necessary, even for nuclear data. Having a good tree for eutherians should enable models to be tested accurately and refined further.

Supplementary Material

Details of the differences between the original hedgehog sequence and the resequenced portions are given in Supplementary Material on the MBE Website, www.molbioevol.org.

Note Added in Proof

After this manuscript was accepted a paper appeared with additional mitochondrial genomes including both a walrus and a pangolin (Arnason et al. 2002). It reports the tree very like figure 4 with hedgehog–gymnure at the base of the eutherian tree, adjacent to the murid rodents. It is only by analyzing the ingroup by itself that the clash between the ingroup tree (fig. 3) and the rooted tree becomes apparent (as in fig. 4). Both figures cannot be correct, and for the reasons outlined in the text we suggest that it is the ingroup tree that is correct, and rooted it looks like figure 5.

Acknowledgments

We thank Dr. Adura Mohd Adnan (Malaysia) and Abby Harrison for the gymnure sample; Cheng Hsi-Chi of Taiwan Endemic Species Research Institute for the shrew and rhinolophid bat tissue samples, and Pdraig Duignan of the Massey University Veterinary School for the sample of the New Zealand fur seal. We also thank Jim Lake for modifications to his triple Markov analysis program, Gambit. The New Zealand Marsden Fund supported this work.

APPENDIX

The Numbers of Evolutionary Trees Without Specifying Subtrees

The number of unrooted binary phylogenetic trees on n taxa is $b(n) = (2n - 5)!!$ (Note that $(2n - 5)!!$ is the product of the first $n - 2$ odd integers, $b(n) = (2n - 5) \times (2n - 7) \times (2n - 9) \dots 5 \times 3 \times 1$.)

Suppose T is a binary phylogenetic tree on the set S of n taxa. If we select a set of $2k - 3$ contiguous edges of T which form a binary subtree T^* , then T/T^* is a set of k subtrees T_1, T_2, \dots, T_k , and these partition S into k subsets. Suppose these subsets contain n_1, n_2, \dots, n_k taxa, respectively.

We can count the number of binary phylogenetic trees on S which induce the same partition of S as follows. Connect the n_i taxa of T_i together in $b(n_i)$ ways, and identify a root on any of the $(2n_i - 3)$ edges, so there are $(2n_i - 3)b(n_i) = b(n_i + 1)$ rooted trees on these n_i taxa. These k rooted subtrees can be linked by their roots in $b(k)$ ways to form a binary phylogenetic tree on S . Hence the number of binary phylogenetic tree on S which induce this partition is $b(k)b(n_1 + 1)b(n_2 + 1) \dots b(n_k + 1)$.

Thus, for example, if $n = 29$, $k = 4$, with $n_1 = 5$, $n_2 = 7$, $n_3 = 8$, and $n_4 = 9$, there are $b(4) \cdot b(6) \cdot b(8) \cdot b(9) \cdot b(10) = 3 \times 105 \times 10,395 \times 135,135 \times 2,027,025 = 8.97 \times 10^{17}$. However, as there are $b(29) = 1.58 \times 10^{35}$ trees on 26 taxa, the probability of getting this partition is 5.68×10^{-18} and for a specific linkage of the four subtrees, this is reduced by a factor of $n(4) = 3$.

LITERATURE CITED

- ADACHI, J., and M. HASEGAWA. 1996. MOLPHY. Computer program published by the authors. Tokyo. <http://bioweb.pasteur.fr/seqanal/interfaces/MolPhy.html>.
- ANDREWS, T. D., and S. EASTEAL. 2000. Evolutionary rate acceleration of cytochrome c oxidase subunit I in simian primates. *J. Mol. Evol.* **50**:562–568.
- ARNASON, U., J. A. ADEGOKE, K. BODIN, E. W. BORN, Y. B. ESA, A. GULLBERG, M. NILSSON, R. V. SHORT, X. XU, and A. JANKE. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl. Acad. Sci. USA* **99**:8151–8156.
- ARNASON, U., A. GULLBERG, A. SCHWEIZER BURGUETE, and A. JANKE. 2001. Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. *Hereditas* **133**:217–228.
- BERTA, A., C. E. RAY, and A. R. WYSS. 1989. Skeleton of the oldest known pinniped, *Enarliarctos mealsi*. *Science* **244**:60–62.
- BUTLER, P. M. 1988. Phylogeny of the insectivores. Pp. 117–141 in M. J. BENTON, ed. *The phylogeny and classification of the tetrapods, Vol. 2. Mammals*. Clarendon Press, Oxford.
- CAO, Y., M. FUJIWARA, M. NIKAIKO, N. OKADA, and M. HASEGAWA. 2000. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene* **259**:149–158.
- CHANG, J. T. 1996. Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. *Math. Biosci.* **134**:189–215.
- CHARLESTON, M. 1994. Factors affecting the performance of phylogenetic methods. Ph.D. thesis, Massey University, Palmerston North.
- COOPER, A., and D. PENNY. 1997. Mass survival of birds across the Cretaceous/Tertiary boundary. *Science* **275**:1109–1113.
- DELISLE, I., and C. STROBECK. 2002. Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. *Mol. Biol. Evol.* **19**:357–361.
- DOUADY, C. J., F. CATZEFELIS, D. J. KAO, M. S. SPRINGER, and M. J. STANHOPE. 2002. Molecular evidence for the monophyly of Tenrecidae (Mammalia). *Mol. Phylogenet. Evol.* **22**:357–363.
- EIZIRIK, E., W. J. MURPHY, and S. J. O'BRIEN. 2001. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* **92**:212–219.
- GOLDMAN, N., J. P. ANDERSON, and A. G. RODRIGO. 2000. Likelihood based tests of topologies in phylogenies. *Syst. Biol.* **49**:652–670.
- HEDGES, S. B., P. H. PARKER, C. G. SIBLEY, and S. KUMAR. 1996. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**:226–229.
- HENDY, M. D., and D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**:297–309.
- HUTCHEON, J. M., J. A. KIRSCH, and J. D. PETTIGREW. 1998. Base-compositional biases and the bat problem III. The questions of microchiropteran monophyly. *Philos. Trans. R. Soc. Lond. B.* **353**:607–617.
- HUTTLEY, G. A., I. B. JAKOBSEN, S. R. WILSON, and S. EASTEAL. 2000. How important is DNA replication for mutagenesis? *Mol. Biol. Evol.* **17**:929–937.
- KARLIN, S., and J. MRÁZEK. 1997. Compositional differences within and between eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* **94**:10227–10232.
- KISHINO, H., J. THORNE, and W. J. BRUNO. 2001. Performance of a divergence time estimation method under a probabilistic model of rate estimation. *Mol. Biol. Evol.* **18**:352–361.
- KRETTEK, A., A. GULLBERG, and U. ARNASON. 1995. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europeus*, and the phylogenetic position of Eulipotyphla. *J. Mol. Evol.* **41**:952–957.
- LAKE, J. 1997. Phylogenetic inference: how much evolutionary history is knowable? *Mol. Biol. Evol.* **14**:213–219.
- LENTO, G. M., R. E. HICKSON, G. K. CHAMBERS, and D. PENNY. 1995. Use of spectral analysis to test hypotheses on the origin of pinnipeds. *Mol. Biol. Evol.* **12**:28–52.
- LIN, Y.-H., and D. PENNY. 2001. Implications for bat evolution from two new complete mitochondrial genomes. *Mol. Biol. Evol.* **18**:684–688.
- LIN, Y.-H., P. J. WADDELL, and D. PENNY. 2002. Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. *Gene* **294**:119–129.
- 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.
- LOU, Z., R. CIFELLI, and Z. KIELAN-JAWOROWSKA. 2001. Dual origin of tribosphenic mammals. *Nature* **409**:53–57.
- MACPHEE, R. D. E., and M. J. NOVACEK. 1993. Definition and relationships of lipotyphla. Pp. 13–31 in F. S. SZALAY, M. J. NOVACEK, and M. C. MCKENNA, eds. *Mammal phylogeny*. Springer-Verlag, New York.
- MADSEN, O., M. SCALLY, C. J. DOUADY, D. J. KAO, R. W. DEBRY, R. ADKINS, H. M. AMRINE, M. J. STANHOPE, W. W. DE JONG, and M. S. SPRINGER. 2001. Molecules reveal parallel adaptive radiations in two major clades of placental mammals. *Nature* **409**:610–614.
- MCKENNA, M. C., and S. K. BELL. 1997. *The classification of mammals: above the species level*. Columbia University Press, New York.
- MOUCHATY, S. K., A. GULLBERG, A. JANKE, and U. ARNASON. 2000a. The phylogenetic position of the Talpidae within eutheria based on analysis of complete mitochondrial sequences. *Mol. Biol. Evol.* **17**:60–67.
- . 2000b. Phylogenetic position of the tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete mitochondrial genome sequence of *Echinops telfairi*. *Zool. Scr.* **29**:307–317.
- MURPHY, W. J., E. EIZIRIK, W. E. JOHNSON, Y. P. ZHANG, O. A. RYDER, and S. J. O'BRIEN. 2001a. Molecular phylogenetics and the origins of placental mammals. *Nature* **409**:614–618.
- MURPHY, W. J., E. EIZIRIK, S. J. O'BRIEN et al. 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**:2348–2351.
- NIKAIKO, M., K. KAWAI, Y. CAO, M. HARADA, S. TOMITA, N. OKADA, and M. HASEGAWA. 2001. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* **53**:508–516.
- OTA, R., P. J. WADDELL, M. HASEGAWA, H. SHIMODAIRA, and H. KISHINO. 2000. Appropriate likelihood ratio tests and

- marginal distributions for evolutionary tree models with constraints on parameters. *Mol. Biol. Evol.* **17**:1417–1424.
- PENNY, D., L. R. FOULDS, and M. D. HENDY. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* **297**:197–200.
- 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.
- PENNY, D., E. E. WATSON, R. E. HICKSON, and P. J. LOCKHART. 1993. Some recent progress with methods for evolutionary trees. *N. Z. J. Bot.* **31**:275–288.
- PHILLIPS, M. J., Y.-H. LIN, G. L. HARRISON, and D. PENNY. 2001. Mitochondrial genomes of a bandicoot and a brush-tail possum confirm the monophyly of australidelphian marsupials. *Proc. R. Soc. Lond., Ser. B.* **268**:1533–1538.
- PUMO, D. E., P. S. FINAMORE, W. R. FRANKE, C. J. PHILLIPS, S. TARZAMI, and D. BALZARANO. 1998. Complete mitochondrial genome of a neotropical fruit bat, *Artibeus jamaicensis*, and a new hypothesis of the relationships of bats to other eutherian mammals. *J. Mol. Evol.* **47**:709–717.
- 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.
- RICH, T. H., P. VICKERS-RICH, A. CONSTANTINE, T. F. FLANNERY, L. KOOL, and N. VAN KLAVEREN. 1997. A tribosphenic mammal from the Mesozoic of Australia. *Science* **278**:1438–1442.
- ROSENBERG, M. S., and S. KUMAR. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. *Proc. Natl. Acad. Sci. USA* **98**:10751–10756.
- 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.
- SPRINGER, M. S., A. BURK, J. R. KAVANAGH, V. G. WADDELL, and M. J. STANHOPE. 1997. The interphotoreceptor retinoid binding protein gene in therian mammals. *Proc. Natl. Acad. Sci. USA* **94**:13754–13759.
- STANHOPE, M. J., V. G. WADDELL, O. MADSEN, W. DE JONG, S. B. HEDGES, G. C. CLEVEN, D. KAO, and M. S. SPRINGER. 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.
- STEEL, M. A., and D. PENNY. 1993. Distributions of tree comparison metrics—some new results. *Syst. Biol.* **42**:126–141.
- STRIMMER, K., and A. RAMBAUT. 2002. Inferring confidence sets of possibly misspecified gene trees. *Proc. R. Soc. Lond. B.* **269**:137–142.
- SWOFFORD, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, 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.
- WADDELL, P. J., H. KISHINO, and R. OTA. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Informatics* **12**:141–154.
- WADDELL, P. J., N. OKADA, and M. HASEGAWA. 1999. Toward resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48**:1–5.
- WADDELL, P. J., and D. PENNY. 1996. Evolutionary trees of apes and humans from DNA sequences. Pp. 53–73 in A. LOCK and C. R. PETERS, eds. *Handbook of human symbolic evolution*. Clarendon Press, Oxford.
- WODZICKI, K. A. 1950. Introduced mammals of New Zealand. Department of Scientific and Industrial Research, Wellington.
- XU, X., and U. ARNASON. 1996. The complete mitochondrial DNA sequence of the great Indian rhinoceros, *Rhinoceros unicornis*, and the phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+Cetacea). *Mol. Biol. Evol.* **13**:1167–1173.

ROSS CROZIER, reviewing editor

Accepted June 5, 2002