

The Complete Mitochondrial DNA Sequence of the Shark *Mustelus manazo*: Evaluating Rooting Contradictions to Living Bony Vertebrates

Ying Cao,*† Peter J. Waddell,* Norihiro Okada,† and Masami Hasegawa*

*The Institute of Statistical Mathematics, Tokyo, Japan; and †Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Japan

A remarkable example of a misleading mitochondrial protein tree is presented, involving ray-finned fishes, coelacanths, lungfishes, and tetrapods, with sea lampreys as an outgroup. In previous molecular phylogenetic studies on the origin of tetrapods, ray-finned fishes have been assumed as an outgroup to the tetrapod/lungfish/coelacanth clade, an assumption supported by morphological evidence. Standard methods of molecular phylogenetics applied to the protein-encoding genes of mitochondria, however, give a bizarre tree in which lamprey groups with lungfish and, therefore, ray-finned fishes are not the outgroup to a tetrapod/lungfish/coelacanth clade. All of the dozens of published phylogenetic methods, including every possible modification to maximum likelihood known to us (such as inclusion of site heterogeneity and exclusion of potentially misleading hydrophobic amino acids), fail to place the ray-finned fishes in a biologically acceptable position. A likely cause of this failure may be the use of an inappropriate outgroup. Accordingly, we have determined the complete mitochondrial DNA sequence from the shark, *Mustelus manazo*, which we have used as an alternative and more proximal outgroup than the lamprey. Using sharks as the outgroup, lungfish appear to be the closest living relative of tetrapods, although the possibility of a lungfish/coelacanth clade being the sister group of tetrapods cannot be excluded.

Introduction

The origin of land vertebrates, or tetrapods, was one of the most important events in the evolution of vertebrates. Sarcopterygians (lobe-finned fishes) have been considered to be closely related to the ancestral tetrapods, because they have bony fins and some of them have lungs for respiration (Romer 1966; Carroll 1988). Living sarcopterygians consist of two groups: coelacanths (only one living species, *Latimeria chalumnae*) and lungfishes (dipnoans). The latter group consists of three genera, which live in the three Gondwanic continents, Africa, South America, and Australia, respectively. The phylogenetic relationship among coelacanths, lungfishes, and tetrapods remains controversial among morphologists as well as among molecular evolutionists. Obviously, a key question is whether the living lobe-finned fishes form a monophyletic group; that will decide whether the taxon Sarcopterygii is acceptable or not in phylogenetic systematics.

Assuming ray-finned fishes as an outgroup, all three possible hypotheses concerning the relationships among coelacanths, lungfishes, and tetrapods have been proposed by morphologists (fig. 1): tree 1, the lungfish/tetrapod clade (e.g., Rosen et al. 1981; Forey 1987); tree 2, the coelacanth/tetrapod clade (e.g., Schultze 1987, 1994); and tree 3, the lungfish/coelacanth clade (e.g., Northcutt 1987; Chang 1991). Thus, it is clear that morphological and paleontological data do not presently resolve the issue with confidence.

Molecular phylogenetics has also been equivocal in this respect. Meyer and Wilson (1990) and Meyer and Dolven (1992), using mitochondrial 12S rRNA and cytochrome *b* sequences, suggested that the lungfish/tet-

rapod clade (tree 1) was real. Gorr, Kleinschmidt, and Fricke (1991), using hemoglobin data, preferred the coelacanth/tetrapod clade (tree 2), which attracted criticism (Forey 1991; Stock et al. 1991). More recently, Yokobori et al. (1994) and Zardoya and Meyer (1996a) preferred the lungfish/coelacanth clade (tree 3) based on mtCOX1 and the cytoplasmic (nuclear) 28S rRNA sequences, respectively. The wheel now seems to have turned full circle; the most recent studies with complete mtDNA sequences (Zardoya and Meyer 1997a; Zardoya et al. 1998) could not distinguish between tree 1 and tree 3 with statistical significance, although tree 2 appeared unlikely. The latest mitochondrial data are a superset (and technically a vastly superior one) of all the previous mtDNA studies. Further, the phylogenetic analyses to date have all used models that are clearly inadequate, in that they ignore major aspects of the evolution of these molecules known to cause potential errors in the analysis (e.g., Lockhart et al. 1994, 1996). Thus, the closest living relative(s) of tetrapods is (are) still not clearly identified (Zardoya and Meyer 1997b; Zardoya et al. 1998).

Furthermore, it should be noted that all of the above-mentioned molecular studies assume that the ray-finned fishes constitute an outgroup to lungfishes, coelacanths, and tetrapods. Although this assumption seems to be strongly supported by the morphological evidence (Carroll 1988; Cloutier and Ahlberg 1996), it needs to be confirmed with molecular evidence.

Although molecular phylogenetics has become a powerful tool for elucidating the evolutionary history of life, there exist many potential pitfalls to inferring the correct tree. Here, in an extraordinary example using the entire protein-coding portion of the mtDNA from diverse vertebrates, we show just how perplexing such cases can be. Using a similar data set, Naylor and Brown (1997, 1998) recently demonstrated that the lancelet is erroneously placed in the Deuterostoma tree 100% of the time in an unweighted parsimony bootstrap analysis.

Key words: molecular phylogeny, misleading tree, origin of tetrapod, outgroup.

Address for correspondence and reprints: Masami Hasegawa, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan. E-mail: hasegawa@ism.ac.jp.

Mol. Biol. Evol. 15(12):1637–1646. 1998

© 1998 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

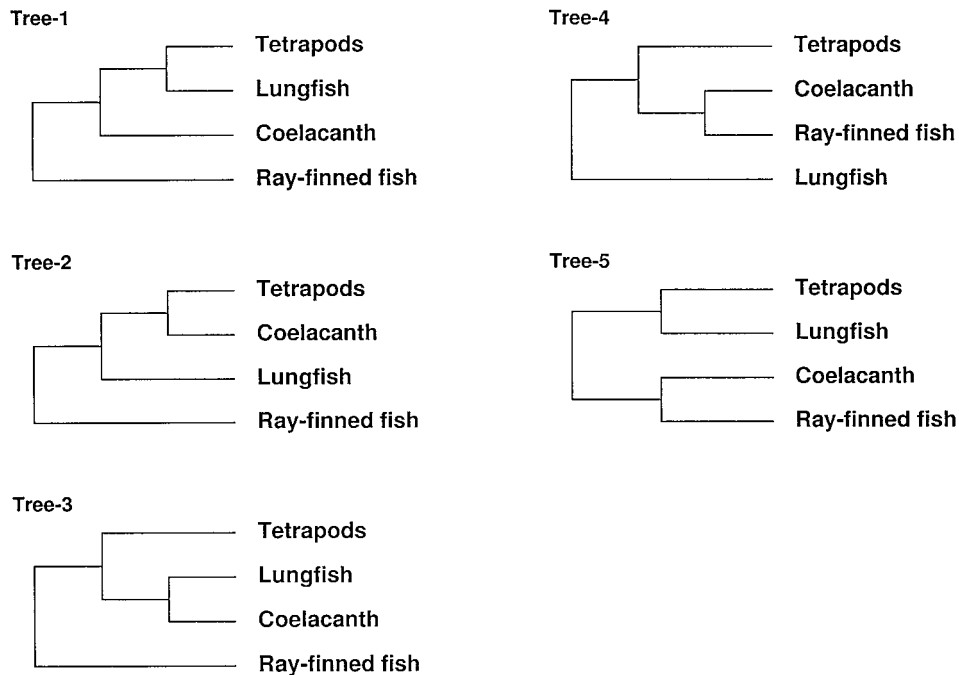


FIG. 1.—Five candidate trees among tetrapods, lungfishes, coelacanths, and ray-finned fishes.

Our own studies have shown another remarkable example of a misleading mitochondrial protein tree among ray-finned fishes, coelacanths (Zardoya and Meyer 1997a), lungfishes (Zardoya and Meyer 1996b), and tetrapods, using sea lampreys (Lee and Kocher 1995) as an outgroup (Zardoya et al. 1998). Nei (1996) obtained a similarly bizarre tree of vertebrates from the mitochondrial protein data (excluding lungfishes and coelacanths). None of the dozens of published phylogenetic methods (see Swofford et al. [1996] for a summary) place the ray-finned fish in a biologically acceptable position with this data set. Only by using mtDNA from a more proximal outgroup, such as sharks, is the problem apparently soluble. The failure when lampreys are used includes every possible modification to maxi-

mum likelihood (ML) we were able to run. Disturbingly, there is presently no objective explanation for this bizarre outcome. This again shows that molecular phylogenetics still has important analytic challenges facing it (Waddell 1995; Lockhart et al. 1996).

Materials and Methods

DNA Extraction and Sequencing

A liver sample was obtained from a gummy shark, *Mustelus manazo*, caught in Tokyo Bay. Total genomic DNA was isolated from fresh liver as described in Cao et al. (1998). A combination of several primers, listed in table 1, were designed based on highly conserved vertebrate mtDNA regions. Localizations and directions

Table 1
Oligonucleotide Primer Sequences Used for Polymerase Chain Reaction

| PCR Primers | Sequence (5' to 3') | Approximate Product Length (bp) |
|---------------------|--|---------------------------------|
| 1. SMet5 | AAGCTWTCGGGCCCATAC | 1,500 |
| 2. SCo13 | CTTAGGGCTGTCCYAACTAT | |
| 3. Sty5 | GYGGGKCTACARYCCACCRCTT | 1,600 |
| 4. SSer3 | GGGGTTTCRAWTCYCYTTTCTTG | |
| 5. SCo15 | CACACCTATGAAGAACCATGC | 1,800 |
| 6. SA163 | GTGCTTRGTGARCAYATTAT | |
| 7. SCo25 | TCAGCARAAGRCKTAYTACATTC | 1,900 |
| 8. SCo33 | GAKCCTCATCARTAGATNGA | |
| 9. Sco35 | CACAGAATTCCACTTTACATCAGAACATCACTTTGGCTTC | 10,000 |
| 10. SNd23 | CACAGAATTCAAATGTAAGAGTAGTTCCCTAGGCCTAAGCTTGA | |

NOTE.—Standard code: R = A/G; Y = C/T; W = A/T; K = G/T; N = G/A/T/C.

Table 2
ProtML Analyses of mtDNA Proteins Based on the mtREV24-F Model with Lampreys as an Outgroup

| Tree Topology | ATP6 | ATP8 | COB | COX1 | COX2 | COX3 |
|---------------------------------|-----------|-----------|-------------|-------------|------------|-----------|
| 1. (((Te,Lu),Co),Ra) | 7.6 ± 5.2 | 1.6 ± 4.1 | 25.1 ± 9.9 | 16.5 ± 9.6 | 12.1 ± 6.2 | 5.3 ± 4.0 |
| 2. (((Te,Co),Lu),Ra) | 7.2 ± 5.4 | 1.6 ± 3.9 | 22.7 ± 10.5 | 16.1 ± 8.1 | 9.6 ± 5.5 | 5.6 ± 3.9 |
| 3. ((Te,(Lu,Co)),Ra) | 3.8 ± 4.1 | 1.6 ± 4.1 | 27.1 ± 10.0 | 8.2 ± 12.4 | 10.7 ± 6.5 | 4.4 ± 4.4 |
| 4. ((Te,(Co,Ra)),Lu) | 5.3 ± 5.9 | 0.1 ± 0.4 | ML | 20.3 ± 8.5 | 1.8 ± 3.4 | ML |
| 5. (((Te,Lu),(Co,Ra)) | 6.8 ± 5.5 | 1.8 ± 4.2 | 13.0 ± 6.2 | 17.9 ± 9.1 | 9.5 ± 6.4 | 2.8 ± 3.3 |
| 6. (Te,(Lu,(Co,Ra))) | 2.8 ± 3.5 | 3.6 ± 3.1 | 7.6 ± 8.1 | 12.6 ± 10.1 | 1.9 ± 8.7 | 0.8 ± 4.6 |
| 7. (Te,((Lu,Co),Ra)) | ML | 3.5 ± 3.4 | 19.4 ± 10.7 | 5.0 ± 12.4 | 1.9 ± 8.8 | 0.9 ± 5.6 |
| 8. (((Te,Lu),Ra),Co) | 6.9 ± 5.4 | 1.8 ± 4.2 | 26.3 ± 9.6 | 6.1 ± 6.5 | 5.0 ± 7.9 | 5.3 ± 4.0 |
| 9. (((Te,Ra),Lu),Co) | 3.8 ± 4.8 | 3.6 ± 3.1 | 28.3 ± 9.7 | 7.9 ± 5.7 | 1.7 ± 7.6 | 3.2 ± 4.4 |
| 10. ((Te,Co),(Lu,Ra)) | 6.8 ± 5.6 | 3.6 ± 3.0 | 22.0 ± 10.7 | 10.1 ± 6.1 | 7.5 ± 6.4 | 3.6 ± 4.9 |
| 11. (Te,(Co,(Lu,Ra))) | 2.4 ± 3.7 | 3.6 ± 3.0 | 18.5 ± 11.2 | 4.5 ± 8.0 | 0.4 ± 8.9 | 1.2 ± 5.6 |
| 12. ((Te,(Lu,Ra)),Co) | 6.3 ± 5.8 | 3.6 ± 3.1 | 24.6 ± 10.7 | ML | 3.3 ± 7.9 | 3.6 ± 4.9 |
| 13. (((Te,Co),Ra),Lu) | 5.2 ± 5.9 | ML | 9.0 ± 7.4 | 17.7 ± 7.4 | ML | 1.6 ± 2.2 |
| 14. (((Te,Ra),Co),Lu) | 2.1 ± 5.2 | 0.1 ± 0.4 | 11.0 ± 6.4 | 20.3 ± 8.5 | 1.0 ± 3.7 | 0.7 ± 3.0 |
| 15. ((Te,Ra),(Lu,Co)) | 3.4 ± 4.3 | 3.5 ± 3.4 | 28.5 ± 9.5 | 9.6 ± 12.0 | 6.4 ± 6.1 | 2.7 ± 4.6 |
| Number of sites | 166 | 32 | 375 | 501 | 214 | 258 |

NOTE.—The highest-likelihood tree among the 15 possible trees among tetrapods (Te), lungfishes (Lu), coelacanths (Co), ray-finned fishes (Ra), and the outgroup is indicated as “ML,” and the differences of log-likelihoods from that of the ML tree are give (±SE). Bootstrap proportion among the 15 possible trees are given in parentheses (RELL method with 10,000 replications).

crete Γ distribution for the site-heterogeneity, and with the wide variety of ML models for nucleotide data (here, first and second codon positions) implemented in MOLPHY (Adachi and Hasegawa 1996b), PAUP* (Swofford 1996), and PHYLIP 3.5 (Felsenstein 1996). The same is true for nucleotide or amino acid parsimony (both weighted and unweighted), plus a wide variety of distance-based methods, such as least-squares, neighbor joining, and split decomposition (based on any available protein distance estimator). Finally, neither linear nor nonlinear invariants (e.g., the Hadamard conjugation) yielded the traditional tree (see Swofford et al. [1996] for descriptions of all of the above phylogenetic methods).

Despite biological expectations, the bizarre tree in figure 3 (tree 4) using lampreys as the outgroup has a log-likelihood (under the ProtML mtREV24-F model) 3 standard errors better than that of any of the “conventional trees” (see “total” in table 2). Further, these traditional trees were never obtained in the 10,000 RELL bootstrap replications we ran on these data. Thus, there is no way the problems can be ascribed to a nonsignificant random error.

It is interesting to note that, while 6 of the 12 individual proteins support tree 4, and most of the remaining genes do not reject this tree, COX1 strongly rejects tree 4, as well as the traditional trees, and supports another bizarre tree, tree 12 (table 2). The message appears to be that, whatever is causing this strange behavior, there is heterogeneity between genes (Cao et al. 1994, 1998). Disturbingly, none of the genes strongly favor the true tree, allowing no easy “ad hoc” fix by concocting reasons to favor these genes.

Taking into account the sites that are unable to change due to functional constraints (we call these invariant sites) is very important. Ignoring this factor alone can lead to an incorrect tree (inconsistency) with all methods or dramatically alter the statistical support for trees (Waddell 1995; Lockhart et al. 1996; Swofford

et al. 1996; Sullivan and Swofford 1997). A robust estimator of the number of invariant sites may be obtained with capture-recapture methods (Sidow, Nguyen, and Speed 1992; Swofford et al. 1996). Here, the removal of this proportion in accordance with the base frequencies at the constant sites (Waddell 1995) massively improved the likelihood of the data (fit of the model to the 3,274-sites data; standard model $\ln L = -55,828.9$, invariant sites model $\ln L = -52,980.4$, with the difference of 2,848.5 being significant; $P < 10^{-100}$). However, it did not change the relative fit of the trees appreciably. For example, tree 4 remained optimal and higher in log-likelihood than did any of the biologically realistic trees (trees 1, 2, and 3) by at least 76.5 log-likelihood units, with a standard error of 21.7 (see “variable sites” in table 3). Thus, it appears that invariant sites are not the cause of, or even a clearcut contributor to, this problem.

Furthermore, the AAML program in PAML (Yang 1997) was applied to the concatenated amino acid sequences taking into account site heterogeneity by the discrete Γ distribution with eight categories. The use of the discrete Γ distribution further improved the likelihood (tree 4, $\ln L = -52,380.9$; tree 6, $\ln L = -52,378.2$). The optimal ML tree has now shifted to a second bizarre tree (tree 6, in which the lamprey lineage is attracted to the tetrapod rather than the lungfish lineage), but the log-likelihood difference between trees 4 and 6 is very minor. Although the log-likelihood differences of traditional trees 1, 2, and 3 from the ML tree are smaller than those for the ProtML analysis with site homogeneity, the differences are still highly significant.

Similarly, the use of LogDeterminant distances (Lake 1994; Lockhart et al. 1994) for either first and second codon positions or amino acids made no difference in the tree, suggesting that the problem is not simply due to base composition or codon biases. The combination of these two factors (the LogDet taking into account invariant-sites) has been seen to be useful in resolving controversial phylogenies based on rRNA

Table 2
Extended

| ND1 | ND2 | ND3 | ND4 | ND4L | ND5 | Total | BP |
|------------|------------|-----------|-------------|-----------|-------------|--------------|--------|
| 10.7 ± 7.9 | 9.5 ± 7.4 | 4.4 ± 4.2 | 21.6 ± 9.6 | 8.0 ± 5.1 | 16.4 ± 14.2 | 101.2 ± 27.3 | 0.0000 |
| 19.2 ± 8.5 | 10.2 ± 8.5 | 7.6 ± 6.6 | 24.7 ± 10.6 | 7.6 ± 5.3 | 22.6 ± 13.4 | 117.2 ± 27.9 | 0.0000 |
| 16.5 ± 9.1 | 10.0 ± 8.5 | 5.5 ± 7.6 | 21.8 ± 11.1 | 8.0 ± 5.1 | 7.8 ± 12.6 | 87.8 ± 29.6 | 0.0000 |
| ML | ML | ML | 4.6 ± 7.8 | ML | 5.6 ± 8.4 | ML | 0.8439 |
| 7.2 ± 5.9 | 3.8 ± 3.0 | 0.3 ± 1.1 | 9.4 ± 6.6 | 4.9 ± 3.2 | 20.7 ± 12.8 | 60.4 ± 22.0 | 0.0000 |
| 5.2 ± 6.7 | 3.8 ± 2.9 | 0.4 ± 1.0 | ML | 4.9 ± 3.2 | 21.1 ± 12.5 | 27.1 ± 22.5 | 0.0637 |
| 10.0 ± 9.1 | 12.3 ± 6.7 | 2.9 ± 5.6 | 2.6 ± 6.3 | 8.2 ± 4.8 | 12.7 ± 11.3 | 41.7 ± 27.2 | 0.0451 |
| 9.9 ± 8.1 | 9.1 ± 6.5 | 4.3 ± 4.2 | 18.7 ± 10.4 | 7.7 ± 5.5 | 11.8 ± 12.7 | 75.3 ± 26.2 | 0.0003 |
| 16.1 ± 9.2 | 7.5 ± 8.2 | 4.7 ± 7.9 | 22.5 ± 11.5 | 7.7 ± 5.5 | 9.6 ± 9.2 | 79.0 ± 26.4 | 0.0003 |
| 20.3 ± 8.1 | 10.0 ± 8.4 | 6.4 ± 7.3 | 22.3 ± 10.7 | 7.6 ± 5.3 | 30.8 ± 11.7 | 113.4 ± 27.0 | 0.0000 |
| 12.0 ± 8.5 | 9.6 ± 8.0 | 4.1 ± 5.3 | 5.2 ± 5.6 | 8.2 ± 4.8 | 27.6 ± 12.1 | 59.7 ± 26.2 | 0.0011 |
| 18.4 ± 8.7 | 8.7 ± 8.6 | 6.4 ± 7.4 | 19.9 ± 11.4 | 7.7 ± 5.5 | 19.9 ± 11.7 | 84.8 ± 27.4 | 0.0002 |
| 9.7 ± 5.9 | 4.2 ± 5.5 | 5.8 ± 5.4 | 17.6 ± 11.4 | 0.5 ± 2.2 | 10.8 ± 7.0 | 44.5 ± 20.7 | 0.0001 |
| 7.9 ± 6.6 | 2.5 ± 6.2 | 3.2 ± 6.8 | 16.4 ± 11.9 | 1.1 ± 1.6 | ML | 28.6 ± 20.9 | 0.0453 |
| 17.3 ± 8.8 | 7.9 ± 7.9 | 4.7 ± 8.0 | 19.9 ± 11.7 | 8.7 ± 4.9 | 10.2 ± 9.3 | 85.3 ± 27.9 | 0.0000 |
| 309 | 309 | 102 | 429 | 91 | 488 | 3,274 | |

(Waddell 1995; Swofford et al. 1996), even when neither factor alone appears able to account for the aberrant behavior. Here, however, the use of this distance estimator (with either protein or nucleotide sequences) combined with various distance-based tree estimation procedures (see earlier) made no qualitative difference (results not shown).

Naylor and Brown (1997, 1998) suggested that isoleucine (I), leucine (L), and valine (V) are the amino

acids responsible for the inconsistency they observed in their example of mtDNA (including the lancelet sequence), and showed that exclusion of these amino acids from the analysis recovers a biologically reasonable tree. Accordingly, it would seem to be no coincidence that the rate of interchange between these three hydrophobic amino acids is very high in mitochondrial proteins (Adachi and Hasegawa 1996a). To test this hypothesis, we reran our analysis by converting all leucines and valines

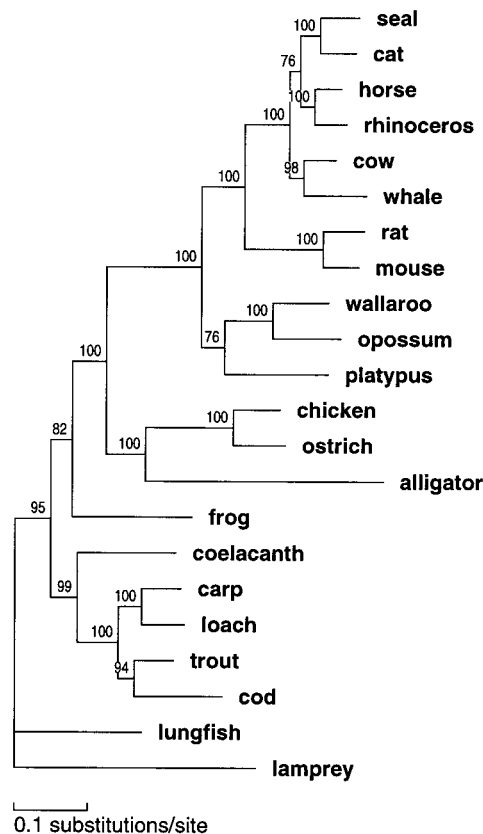


FIG. 3.—Evolutionary trees of vertebrates with lampreys as an outgroup estimated from mitochondrial proteins using ProtML with the mtREV24-F model (Adachi and Hasegawa 1996b). For each internal edge, the bootstrap probability (%) (after fixing the relationships within subtrees attached to that branch) is shown. Local bootstrap probability (Adachi and Hasegawa 1996b) was estimated with the RELL method with 10,000 replications. The horizontal length of each branch is proportional to the estimated number of amino acid substitutions.

Table 3
ProtML and AAML Analyses of Concatenated mtDNA-Encoded Proteins, and NucML and BASEML Analyses of Concatenated mt-tRNAs with Lampreys as an Outgroup

| TREE | 12 PROTEINS | | | | | | | |
|----------------------------|----------------------|----------------------|---------------------------|--------------------------|--------------------------|---|------------------|----------------------------|
| | ProtML | | | | AAML Gamma (3,274) | NucML First and Second Positions (6,548) | tRNAs | |
| | All Sites (3,274) | I = L = V (3,274) | Variable Sites (2,221) | Perfect Sites (1,438) | | | NucML (1,233) | BASEML Gamma (1,206) |
| 1. (((Te,Lu),Co),Ra) ... | 119.8 ± 22.8 | 110.4 ± 22.5 | 77.5 ± 17.5 | 10.9 ± 7.7 | 33.0 ± 12.3 | 138.8 ± 24.7 | 5.4 ± 8.9 | 0.9 ± 4.1 |
| 2. (((Te,Co),Lu),Ra) ... | 140.3 ± 25.9 | 122.8 ± 25.6 | 90.1 ± 20.3 | 10.9 ± 7.7 | 44.9 ± 13.9 | 164.2 ± 27.2 | 6.3 ± 10.1 | 3.1 ± 4.9 |
| 3. ((Te,(Lu,Co)),Ra) ... | 117.3 ± 28.0 | 104.1 ± 27.3 | 76.5 ± 21.7 | 8.8 ± 8.2 | 35.0 ± 15.1 | 136.2 ± 29.5 | 2.1 ± 10.6 | 1.9 ± 5.2 |
| 4. ((Te,(Co,Ra)),Lu) ... | ML | ML | ML | 0.4 ± 8.0 | 2.7 ± 9.3 | ML | ML | ML |
| 5. ((Te,Lu),(Co,Ra)) ... | 61.1 ± 14.4 | 59.1 ± 14.9 | 37.2 ± 11.0 | 7.0 ± 6.1 | 13.4 ± 7.3 | 75.6 ± 15.5 | 4.9 ± 4.7 | 0.2 ± 2.7 |
| 6. (Te,(Lu,(Co,Ra))) ... | 30.5 ± 18.7 | 28.8 ± 19.2 | 17.0 ± 14.5 | ML | ML | 39.2 ± 19.9 | 3.0 ± 5.4 | 0.8 ± 2.3 |
| 7. (Te,((Lu,Co),Ra)) ... | 61.9 ± 27.4 | 47.6 ± 26.7 | 35.8 ± 21.3 | 1.0 ± 4.8 | 12.1 ± 10.8 | 79.4 ± 29.3 | 0.3 ± 9.5 | 1.2 ± 5.0 |
| 8. (((Te,Lu),Ra),Co) ... | 96.6 ± 24.2 | 87.7 ± 24.0 | 64.2 ± 18.5 | 8.0 ± 8.4 | 28.3 ± 12.8 | 115.3 ± 26.1 | 5.5 ± 8.4 | 0.6 ± 4.2 |
| 9. (((Te,Ra),Lu),Co) ... | 103.4 ± 28.1 | 88.1 ± 27.8 | 67.4 ± 21.9 | 8.0 ± 8.4 | 32.7 ± 15.7 | 122.7 ± 30.0 | 4.1 ± 9.5 | 3.1 ± 4.6 |
| 10. ((Te,Co),(Lu,Ra)) ... | 136.2 ± 26.4 | 116.1 ± 26.3 | 87.4 ± 20.6 | 9.8 ± 6.9 | 46.2 ± 13.4 | 155.9 ± 28.0 | 9.2 ± 9.4 | 3.2 ± 4.9 |
| 11. (Te,(Co,(Lu,Ra))) ... | 78.5 ± 26.6 | 62.0 ± 26.2 | 47.2 ± 20.7 | 3.0 ± 3.9 | 22.5 ± 8.9 | 89.4 ± 29.1 | 5.9 ± 8.9 | 2.6 ± 4.6 |
| 12. (((Te,(Lu,Ra)),Co) ... | 111.3 ± 28.3 | 93.5 ± 28.3 | 74.1 ± 21.9 | 6.9 ± 7.7 | 39.8 ± 14.3 | 126.7 ± 30.3 | 4.8 ± 10.0 | 2.6 ± 5.0 |
| 13. (((Te,Co),Ra),Lu) ... | 58.5 ± 17.3 | 49.6 ± 17.0 | 40.9 ± 14.2 | 4.2 ± 9.2 | 30.9 ± 14.4 | 67.1 ± 17.9 | 4.0 ± 6.9 | 1.9 ± 3.9 |
| 14. (((Te,Ra),Co),Lu) ... | 42.5 ± 19.3 | 34.3 ± 19.0 | 30.1 ± 15.9 | 4.2 ± 9.2 | 23.2 ± 15.9 | 48.7 ± 20.6 | 2.6 ± 7.1 | 2.3 ± 3.6 |
| 15. ((Te,Ra),(Lu,Co)) ... | 108.5 ± 27.7 | 93.5 ± 27.3 | 69.5 ± 21.7 | 8.8 ± 8.2 | 32.0 ± 15.6 | 131.6 ± 29.3 | 3.6 ± 9.6 | 2.4 ± 4.9 |

NOTE.—The α/β ratio of the HKY85 model in the mt-tRNA analysis was estimated to be 7.16.

into isoleucine (thus, the analyses would not discriminate among these three amino acids). The result was no significant difference in any of the many different types of phylogenetic analysis (see “I = L = V” in table 3). Thus, Naylor and Brown’s (1997, 1998) recommendation does not improve the situation, despite mtDNA sequences being similar to those they studied, suggesting that the problem must lie elsewhere in this case.

Finally, we considered a method based on the “perfect-sites” analysis (Waddell 1995), which uses a capture-recapture philosophy to identify, then remove, sites experiencing a higher rate of change. Here, all positions showing any change in either the monophyletic group of tetrapods or the monophyletic group of ray-finned fishes were removed, leaving 1,438 sites, many of them unvaried. This ensures an unbiased set generally of the slowest-evolving sites, which are expected to be most robust to misspecification of an ML model. Tree 6 is favored as in the AAML analysis, with tree 4 being practically equally good, while the biologically realistic trees are still all unlikely. This type of site screening dramatically reduces the amount of data and its apparent resolving power. The shift in ML to the second unusual tree, tree 6, completes a trend in its ascendancy over tree 4 as the selection of sites and model becomes increasingly stringent. The other methods of analysis (not shown) follow a similar shift in preference toward tree 6, and all of the biologically realistic trees remain highly improbable.

Breaking Up the Long Edge to Lampreys

A denser sampling of Agnatha might help resolve the situation if the sea lamprey sequence was aberrant. In order to check this, we reanalyzed ND1 and ND2 with the ProtML, including data from two other agnathan species, the river lamprey *Lampetra fluviatilis* and the hagfish *Myxine glutinosa*, sequenced by Delarbre et

al. (1997), as an outgroup in addition to the sea lamprey. However, the result remained almost unchanged from that shown in table 2; i.e., tree 4 remains the ML tree for both proteins, and trees 1, 2, and 3 have log-likelihood values lower than does tree 4 by 10.7 ± 7.9 (7.1 ± 8.1), 19.2 ± 8.5 (7.5 ± 8.8), and 16.5 ± 9.1 (7.4 ± 8.8), respectively, for ND1 (ND2). In these trees, the hagfish and the lamprey were very distantly related, although there is disagreement as to whether they constitute a monophyletic group or a paraphyletic group (e.g., Stock and Whitt 1992; Rasmussen, Janke, and Arnason 1998). Both the hagfish and the lamprey joined the jawed vertebrate tree at a similar strange locality, showing that the lamprey sequence is not unusual in this regard (and eliminating possible reasons, such as gross sequencing errors or a chimeric sequence).

Analyzing tRNAs

It might be expected that mt-tRNA sequences would give a reliable tree for the early vertebrate branchings due to their fairly slow rate of change (Kumazawa and Nishida 1993, 1995a, 1995b). However, the unusual tree 4 is the ML tree for mt-tRNAs too, while the log-likelihood differences among any possible trees are very minor (table 3). The differences are even more minor when the BASEML program in PAML is applied with the discrete Γ distribution (the number of nucleotide sites is reduced in the BASEML analysis, because this program excludes gap sites automatically), suggesting that the mt-tRNAs do not have sufficient reliable information to resolve this problem.

Of course, there remains a possibility that tree 4 in figure 1 (or tree 6) is indeed correct, but this seems unlikely from both the morphological evidence (Carroll 1988; Cloutier and Ahlberg 1996) and the analysis presented below. The possibility that the living Agnatha are misclassified and are really very degenerate forms of

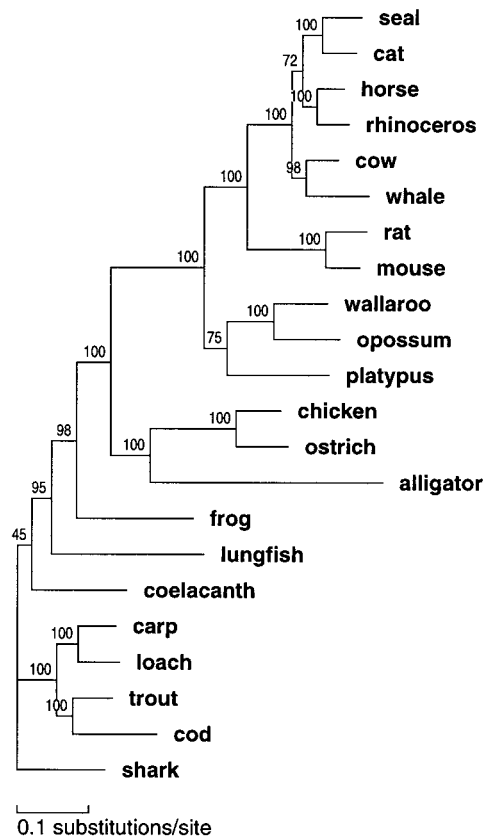


FIG. 4.—Evolutionary trees of vertebrates with sharks as an outgroup estimated from mitochondrial proteins using ProtML with the mtREV24-F model (Adachi and Hasegawa 1996b). This tree (tree 1) is not the highest-likelihood tree under the model, but the log-likelihood difference from the best tree (tree 5), which has no support from morphological evidence, is only 0.8 ± 17.9 .

advanced jawed bony fish rather than relicts of the early jawless fish radiation likewise seems highly improbable.

Resolving the Issue with the Shark Sequence

To help solve this puzzle, we sequenced the whole mtDNA of gummy shark, *M. manazo*. This taxon is closer to the ingroup species than is the lamprey, and so it is expected to be a better outgroup. The sequence has been deposited in the DDBJ/EMBL/GenBank database (accession number AB015962). This sequence is 16,707 bp long, and all of the typical vertebrate protein, rRNA, and tRNA genes are present, while the gene arrangement is identical to that in the sequenced ray-finned fish mtDNAs.

By using sharks instead of lampreys as the outgroup, a traditional tree (tree 1) showing ray-finned fishes as the sister taxa to the remaining bony vertebrates is optimal with some methods. While a nontraditional tree (tree 5) is optimal using ProtML with the site-homogeneous mtREV24-F model, its log-likelihood value is practically indistinguishable from that of a traditional tree (tree 1; fig. 4) (with a difference of only 0.8 ± 17.9 ; table 4). Trees 2, 3, and 4 have log-likelihoods substantially worse than that of tree 5 by 33.3 ± 23.6 , 20.6 ± 24.6 , and 22.0 ± 15.8 , respectively. With the AAML program, tree 1 turned out to be the ML tree, although the log-likelihood difference between trees 1 and 5 is still minor. All other phylogenetic methods give similar

results, suggesting the possibility that the shark is indeed behaving as a reasonable outgroup (to the extent that it does not significantly favor any bizarre tree over the reasonable trees; results not shown).

The recovery of tree 1 as likely (although not unique) is significant, as there has been some disagreement among morphologists on this point. Specifically, ray-finned fishes are definitely less closely related to tetrapods than are lobe-finned fishes (for review see, e.g., Forey 1998). Tree 5 is not one of the trees the morphologists thought might be correct, so, based on a Bayesian argument, the data have some resolving power to suggest that ray-finned fishes are an outgroup to all other bony vertebrates. However, the small differences separating these trees suggest that the time interval for these divergences may be shorter than is commonly accepted. The BASEML and PAUP* ML analyses of mt-tRNAs with a Γ distribution of site rates with sharks as the outgroup again suggest that the data do not contain sufficient information to resolve the phylogenetic problem at hand.

Finally, analyses using both sharks and lampreys clearly show the very distinct characteristics of these two sequences (trees not shown). The edge leading to lampreys is much longer than the edge leading to sharks. These two taxa do not appear together on the tree, and lampreys show a tendency to join near frogs or the lin-

Table 4
ProtML and AAML Analyses of Concatenated mtDNA-Encoded Proteins, and NucML and BASEML Analyses of Concatenated mt-tRNAs with Sharks as an Outgroup

| TREE | 12 PROTEINS | | | | tRNAs | |
|-------------------------------|----------------------|----------------------|---------------------------|--------------------------|------------------|----------------------------|
| | ProtML | | | AAML Gamma (3,274) | NucML (1,233) | BASEML Gamma (1,206) |
| | All Sites (3,274) | I = L = V (3,274) | Variable Sites (2,184) | | | |
| 1. (((Te,Lu),Co),Ra) | 0.8 ± 17.9 | 1.3 ± 17.4 | 1.5 ± 14.9 | ML | 14.9 ± 10.8 | 7.8 ± 6.4 |
| 2. (((Te,Co),Lu),Ra) | 33.3 ± 23.6 | 23.8 ± 23.0 | 23.4 ± 19.4 | 13.3 ± 6.9 | 22.5 ± 10.1 | 11.2 ± 5.8 |
| 3. ((Te,(Lu,Co)),Ra) | 20.6 ± 24.6 | 12.6 ± 23.8 | 15.9 ± 20.1 | 6.2 ± 8.5 | 15.0 ± 11.5 | 9.1 ± 6.5 |
| 4. ((Te,(Co,Ra)),Lu) | 22.0 ± 15.8 | 24.0 ± 15.8 | 14.1 ± 12.8 | 9.6 ± 14.8 | 15.4 ± 11.6 | 6.8 ± 7.3 |
| 5. ((Te,Lu),(Co,Ra)) | ML | ML | ML | 1.9 ± 10.9 | 12.5 ± 11.4 | 5.4 ± 7.3 |
| 6. (Te,(Lu,(Co,Ra))) | 35.1 ± 14.1 | 33.0 ± 14.4 | 23.8 ± 11.2 | 7.2 ± 15.6 | 12.3 ± 11.7 | 5.0 ± 7.4 |
| 7. (Te,((Lu,Co),Ra)) | 38.6 ± 22.0 | 27.2 ± 21.3 | 27.8 ± 17.5 | 6.5 ± 15.3 | 6.7 ± 12.3 | 4.9 ± 7.2 |
| 8. (((Te,Lu),Ra),Co) | 17.1 ± 15.3 | 16.6 ± 14.7 | 13.0 ± 12.6 | 8.8 ± 9.2 | 4.3 ± 7.2 | 2.3 ± 4.6 |
| 9. (((Te,Ra),Lu),Co) | 51.5 ± 24.7 | 49.4 ± 23.7 | 35.7 ± 20.7 | 17.1 ± 15.0 | 6.4 ± 6.5 | 3.6 ± 3.9 |
| 10. ((Te,Co),(Lu,Ra)) | 60.5 ± 23.2 | 49.3 ± 22.5 | 44.4 ± 18.8 | 28.0 ± 11.7 | 14.3 ± 6.4 | 6.2 ± 3.9 |
| 11. (Te,(Co,(Lu,Ra))) | 59.4 ± 20.9 | 49.2 ± 20.1 | 41.6 ± 16.6 | 17.1 ± 15.1 | 6.8 ± 8.8 | 3.2 ± 5.4 |
| 12. ((Te,(Lu,Ra)),Co) | 63.2 ± 22.7 | 58.6 ± 21.9 | 47.4 ± 18.5 | 25.8 ± 13.0 | ML | ML |
| 13. (((Te,Co),Ra),Lu) | 33.8 ± 22.6 | 27.6 ± 22.0 | 23.7 ± 18.7 | 20.8 ± 12.8 | 20.1 ± 10.2 | 10.2 ± 6.1 |
| 14. (((Te,Ra),Co),Lu) | 31.7 ± 24.3 | 31.3 ± 23.4 | 21.2 ± 20.6 | 13.0 ± 15.7 | 14.7 ± 10.3 | 8.2 ± 6.1 |
| 15. ((Te,Ra),Lu,Co) | 39.0 ± 25.8 | 32.5 ± 24.8 | 27.9 ± 21.4 | 10.9 ± 15.1 | 11.4 ± 10.9 | 7.4 ± 6.3 |

edge leading to amniotes (standard ProtML tree), and sometimes even near the lineage leading to birds! The shark, in contrast, joins on the edge leading to ray-finned fishes and coelacanths.

Concerning the relationship among coelacanths, lungfishes, and tetrapods, our analysis, despite uncertainties, tends to prefer the lungfish/tetrapod grouping, although we could not exclude the other possibilities. Our analyses are more extensive and sophisticated than those of Zardoya and Meyer (1997a) and Zardoya et al. (1998), but we have each reached the same conclusion: nobody can offer statistically sound support resolving these relationships. Even using sharks as an outgroup, we could not discriminate between tree 1 and tree 5. While it is accepted from paleontology that the divergence time between the living lobe-finned fish lineages and the tetrapod lineage occurred in as little as 15 MYA (Carroll 1997; Zardoya and Meyer 1997b), the similar fit of trees 1 and 5 may suggest that the ray-finned fish did not branch off much earlier, something that is less clear from the fossil record.

Conclusions

This analysis clearly suggests that something very unusual has occurred in the lineage leading not just to the sea lamprey mtDNA, but to that of other agnathans as well. We seem to have eliminated the commonly suggested causes (unequal site rates, certain amino acids being unreliable, long edges attracting, unequal base composition), leaving presently only enigmatic, but imprecisely specified possibilities, such as shifts in the functional roles of amino acid sites, or natural selection fixing substitutions in a highly peculiar way. These ghostly apparitions at the edge of current phylogenetic thought need to be explored more fully, and the evolution of lampreys appears to provide an ideal case study. A complete mtDNA sequence of the other living agna-

than, the hagfish, could be especially illuminating in this regard.

Indeed, since submission of this paper for review, a hagfish sequence has been published (Rasmussen, Janke, and Arnason 1998). While the sequence is a valuable contribution, we are very surprised at the authors' conclusion: that Agnatha really do root on the edge leading to amniotes, hence making all extant jawed fish monophyletic. As we have shown with analyses of the new shark sequence, we do not see any good evidence to support such an extraordinary claim. While it is true that the shark does not statistically exclude the rooting point for the lamprey, it is also true, and unexplainable, that the long-edged lamprey does significantly exclude the rooting point for the shark. Such a situation does not make sense modelwise, leading only to a prediction that the model is not adequate for reliably locating the lamprey sequence.

A denser sampling of not only Agnatha, but also Chondrichthyes, Dipnoi, Amphibia, and Reptilia, may improve the phylogenetic estimate from the mitochondrial proteins by helping methods like ML and parsimony to reconstruct likely ancestral states more reliably. For example, even though strong convergent evolution has clearly occurred in the lysozymes of ruminant artiodactyls and leaf-eating monkeys, causing them to be erroneously combined in the phylogenetic analyses of small numbers of species (Stewart, Schilling, and Wilson 1987), a reasonable tree is sometimes obtained when the number of sampled species is increased, in spite of the presence of such misleading signals (Adachi and Hasegawa 1996b).

To intractively diagnose the cause of an erroneous tree as "long edges attracting" is often not helpful (Waddell 1995). Long-edge attraction is a symptom of an underlying problem, that of more parallelisms and convergences than expected under the model used to

reconstruct the tree. This much is obvious. While long edges are sometimes apparent in such cases, they need not be, and they tell us little of the nature of the problem. What is really required is a detailed diagnosis. In the present case, the cause has eluded us. We have found no evidence for unequal site rates, varying base composition, a misspecified transition matrix (e.g., changes between I, L, and V being much more common than anticipated), or any combination of these. In this example, the agnathans represent relatively long lineages, but the significant results favoring the wrong tree show that a strong model-violating bias is in operation. Long edges are, at best, a symptom and not a cause. Supporting this view are two important facts: (1) the second-best ProtML tree involves the joining of a short internal edge (to tetrapods) and a long external edge (to lampreys); and (2) as more stringent models are used, better taking into account multiple hits, this second-best tree actually becomes the best tree. This is strange indeed, as this new rooting point is no closer to the expected biological root, near shark and ray-finned fish. There must be a biological reason for this behavior. Until it can be identified and removed or accommodated, molecular phylogenetics is left with this worrying puzzle and another cautionary tale.

To conclude, the recovery of an acceptable tree by replacing lampreys with sharks demonstrates a number of important points: (1) the importance of choosing an appropriate outgroup in phylogenetic estimation; (2) that all methods of phylogenetic analysis can be severely misleading, including ML (Waddell 1995); (3) that a conciliation of results from different algorithms is *not* necessarily good evidence for the robustness or reliability of an analysis; and (4) that phylogenetics and molecular structural biology are still in their infancy when it comes to explaining or working around abnormal modes of sequence evolution.

Acknowledgments

This paper constitutes a part of the research done by Y. C. in partial fulfillment of a Ph.D. degree at Tokyo Institute of Technology. We thank Axel Meyer and Rafael Zardoya for helpful discussions in the early stages of this project, and T. Kamaishi for technical help. We also thank Ziheng Yang for help in using PAML, and anonymous reviewers for helpful comments that improved the manuscript. This work was supported by grants from the Ministry of Education, Science, Sports and Culture of Japan (M.H., N.O., and P.J.W.), JSPS (P.J.W.), and the Marsden Fund (P.J.W.).

LITERATURE CITED

- ADACHI, J., and M. HASEGAWA. 1996a. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* **42**:459–468.
- . 1996b. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**:1–150.
- CAO, Y. 1998. Molecular phylogeny and evolution of vertebrates. Ph.D. thesis, Tokyo Institute of Technology.
- CAO, Y., J. ADACHI, A. JANKE, S. PÄÄBO, and M. HASEGAWA. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *J. Mol. Evol.* **39**:519–527.
- CAO, Y., A. JANKE, P. J. WADDELL, M. WESTERMAN, O. TAKENAKA, S. MURATA, N. OKADA, S. PÄÄBO, and M. HASEGAWA. 1998. Conflict amongst individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *J. Mol. Evol.* **47**:307–322.
- CARROLL, R. L. 1988. Vertebrate paleontology and evolution. Freeman, New York.
- CARROLL, R. L. 1997. Patterns and processes of vertebrate evolution. Cambridge University Press, Cambridge.
- CHANG, M.-M. 1991. “Rhipidistian”, dipnoans, and tetrapods. Pp. 3–28 in H.-P. SCHULTZE and L. TRUEB, eds. *Origins of the higher groups of tetrapods—controversy and consensus*. Cornell University Press, Ithaca, N.Y.
- CLOUTIER, R., and P. E. AHLBERG. 1996. Morphology, characters, and the interrelationships of basal sarcopterygians. Pp. 445–479 in M. L. J. STIASNY, L. R. PARENTI, and G. D. JOHNSON, eds. *Interrelationships of fishes*. Academic Press, San Diego.
- DELARBRE, C., V. BARRIEL, S. TILLIER, P. JANVIER, and G. GACHELIN. 1997. The main features of the Craniate mitochondrial DNA between the ND1 and the COI genes were established in the common ancestor with the lancelet. *Mol. Biol. Evol.* **14**:807–813.
- ELSENSTEIN, J. 1996. PHYLIP (phylogeny inference package). Version 3.572. Distributed by the author, Department of Genetics, University of Washington, Seattle.
- FOREY, P. L. 1987. Relationships of lungfishes. Pp. 75–91 in W. E. BEMIS, W. W. BURGGREN, and N. KEMP, eds. *The biology and evolution of lungfishes*. Alan R. Liss, New York.
- . 1991. Blood lines of the coelacanth. *Nature* **351**:347–348.
- . 1998. *History of the coelacanth fishes*. Chapman and Hall, London.
- GORR, T., T. KLEINSCHMIDT, and H. FRICKE. 1991. Close tetrapod relationships of the coelacanth *Latimeria* indicated by haemoglobin sequences. *Nature* **351**:394–397.
- HÄRLID, A., A. JANKE, and U. ARNASON. 1997. The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. *Mol. Biol. Evol.* **14**:754–761.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- JANKE, A., and U. ARNASON. 1997. The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent Archosauria (birds and crocodiles). *Mol. Biol. Evol.* **14**:1266–1272.
- JANKE, A., N. J. GEMMELL, G. FELDMAIER-FUCHS, A. VON HAESLER, and S. PÄÄBO. 1996. The complete mitochondrial genome of a monotreme, the platypus (*Ornithorhynchus anatinus*). *J. Mol. Evol.* **42**:153–159.
- JANKE, A., X. XU, and U. ARNASON. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia and Eutheria. *Proc. Natl. Acad. Sci. USA* **94**:1276–1281.
- JOHANSEN, S., and I. BAKKE. 1996. The complete mitochondrial DNA sequence of Atlantic cod (*Gadus morhua*): relevance to taxonomic studies among codfishes. *Mol. Mar. Biol. Biotechnol.* **5**:203–214.

- KUMAZAWA, Y., and M. NISHIDA. 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *J. Mol. Evol.* **37**:380–398.
- . 1995a. Variation in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. *Mol. Biol. Evol.* **12**:759–772.
- . 1995b. Phylogenetic utility of mitochondrial transfer RNA genes for deep divergence in animals. Pp. 23–35 in M. NEI and N. TAKAHATA, eds. *Current topics on molecular evolution—proceedings of the U.S.–Japan workshop*. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, and Graduate University for Advanced Studies, Japan.
- LAKE, J. A. 1994. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc. Natl. Acad. Sci. USA* **91**:1455–1459.
- LEE, W. J., and T. D. KOCHER. 1995. Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics* **139**:873–887.
- LOCKHART, P. J., A. W. D. LARKUM, M. A. STEEL, P. J. WADDELL, and D. PENNY. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proc. Natl. Acad. Sci. USA* **93**:1930–1934.
- 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.
- LOPEZ, J. V., M. CULVER, S. CEVARIO, and S. J. O'BRIEN. 1996. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA repeat (Numt) in the nuclear genome. *Genomics* **33**:229–246.
- MEYER, A., and S. I. DOLVEN. 1992. Molecules, fossils, and the origin of tetrapods. *J. Mol. Evol.* **35**:102–113.
- MEYER, A., and A. C. WILSON. 1990. Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. *J. Mol. Evol.* **31**:359–364.
- NAYLOR, G. J. P., and W. M. BROWN. 1997. Structural biology and phylogenetic estimation. *Nature* **388**:527–528.
- . 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* **47**:61–76.
- NEI, M. 1996. Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.* **30**:371–403.
- NORTHCUTT, R. G. 1987. Lungfish neural characters and their bearing on sarcopterygian phylogeny. Pp. 277–297 in W. E. BEMIS, W. W. BURGGREN, and N. KEMP, eds. *The biology and evolution of lungfishes*. Alan R. Liss, New York.
- RASMUSSEN, A. S. G., A. JANKE, and U. ARNASON. 1998. The mitochondrial DNA molecule of the hagfish (*Myxine glutinosa*) and vertebrate phylogeny. *J. Mol. Evol.* **46**:382–388.
- ROMER, A. S. 1966. *Vertebrate paleontology*. University of Chicago Press, Chicago.
- ROSEN, D. E., P. L. FOREY, B. G. GARDINER, and C. PATTERSON. 1981. Lungfishes, tetrapods, paleontology and pleisiomorphy. *Bull. Am. Mus. Nat. Hist.* **167**:159–276.
- SCHULTZE, H.-P. 1987. Dipnoans as sarcopterygians. Pp. 39–74 in W. E. BEMIS, W. W. BURGGREN, and N. KEMP, eds. *The biology and evolution of lungfishes*. Alan R. Liss, New York.
- . 1994. Comparison of hypotheses on the relationships of sarcopterygians. *Syst. Biol.* **43**:155–173.
- SIDOW, A., T. NGUYEN, and T. P. SPEED. 1992. Estimating the fraction of invariable codons with a capture-recapture method. *J. Mol. Evol.* **35**:253–260.
- STEWART, C.-B., J. W. SCHILLING, and A. C. WILSON. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* **330**:401–404.
- STOCK, D. W., D. L. SWOFFORD, P. M. SHARP, A. T. LLOYD, D. G. HIGGINS, A. MEYER, and A. C. WILSON. 1991. Coelacanth relationships. *Nature* **353**:217–219.
- STOCK, D. W., and G. S. WHITT. 1992. Evidence from 18S ribosomal RNA sequences that lampreys and hagfishes form a natural group. *Science* **257**:787–789.
- STRIMMER, K., and A. VON HAESLER. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**:964–969.
- SULLIVAN, J., and D. L. SWOFFORD. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J. Mamm. Evol.* **4**:77–86.
- SWOFFORD, D. L. 1996. PAUP*. National Museum of Natural History, Smithsonian Institution, Washington, D.C.
- SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 in D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. *Molecular systematics*. 2nd edition. Sinauer, Sunderland, Mass.
- WADDELL, P. J. 1995. *Statistical methods of phylogenetic analysis: including Hadamard conjugations, LogDet transforms, and maximum likelihood*. Ph.D. dissertation, Massey University.
- XU, X., A. JANKE, and U. ARNASON. 1996. The complete mitochondrial DNA sequences of the greater Indian rhinoceros, *Rhinoceros unicornis*, and the phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+ Cetacea). *Mol. Biol. Evol.* **13**:1167–1173.
- YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**:555–556.
- YOKOBORI, S., M. HASEGAWA, T. UEDA, N. OKADA, K. NISHIKAWA, and K. WATANABE. 1994. Relationship among coelacanths, lungfishes and tetrapods; a phylogenetic analysis based on mitochondrial cytochrome oxidase I gene sequences. *J. Mol. Evol.* **38**:602–609.
- ZARDOYA, R., Y. CAO, M. HASEGAWA, and A. MEYER. 1998. Searching for the closest living relative(s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data. *Mol. Biol. Evol.* **15**:506–517.
- ZARDOYA, R., A. GARRIDO-PERTIERRA, and J. M. BAUTISTA. 1995. The complete nucleotide sequence of mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. *J. Mol. Evol.* **41**:942–951.
- ZARDOYA, R., and A. MEYER. 1996a. The origin of tetrapods based on 28S ribosomal RNA sequences. *Proc. Natl. Acad. Sci. USA* **93**:5449–5454.
- . 1996b. The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. *Genetics* **142**:1249–1263.
- . 1997a. The complete DNA sequence of the mitochondrial genome of a “living fossil”, the coelacanth (*Latimeria chalumnae*). *Genetics* **146**:995–1010.
- . 1997b. Molecular phylogenetic information on the identity of the closest living relative(s) of land vertebrates. *Naturwissenschaften* **84**:389–397.

MITIKO GO, reviewing editor

Accepted August 28, 1998