

The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths

Jun G. Inoue^{a,*}, Masaki Miya^b, Byrappa Venkatesh^c, Mutsumi Nishida^a

^a*Ocean Research Institute, The University of Tokyo, 1-15-1 Minamidai, Nakano-ku, Tokyo 164-8639, Japan*

^b*Department of Zoology, Natural History Museum and Institute, Chiba, 955-2 Aoba-cho, Chuo-ku, Chiba 260-8682, Japan*

^c*Institute of Molecular and Cell Biology, Proteos Building, 61 Biopolis Drive, Singapore 138673, Singapore*

Received 19 October 2004; accepted 6 January 2005

Available online 2 March 2005

Received by T. Sekiya

Abstract

We determined the whole mitochondrial genome sequence for Indonesian coelacanth *Latimeria menadoensis*. The genome content and organization were identical to that of typical vertebrates including Comoran coelacanth, *Latimeria chalumnae*. The overall nucleotide differences between the two species (excluding the control region) was 4.28%. The divergence time between the two species was estimated using whole mitochondrial genome data from the two coelacanths and 26 actinopterygians that represent major actinopterygian lineages plus an outgroup. Partitioned Bayesian analyses were conducted with the two data sets that comprised concatenated amino acid sequences from 12 protein-coding genes (excluding ND6 gene) and concatenated nucleotide sequences from 12 protein-coding genes (without 3rd codon positions), 22 transfer RNA genes, and two ribosomal RNA genes. The molecular clock analysis was also conducted with the concatenated amino acid sequences from the 12 protein-coding genes after removing faster or more slowly evolving sequences. Using the sarcopterygian–actinopterygian split as a calibration point (450 Mya), divergence time estimation between *L. menadoensis* and *L. chalumnae* fell in the range of 40–30 Mya, which is much older than those of the previous studies (<6.3 Mya). Assuming that the most recent ancestor of *Latimeria* was distributed continuously along the deep coasts of Africa through Eurasia, our estimate is in agreement with the hypothesis that the collision of India with Eurasia (50 Mya) and the subsequent siltation caused by the formation of major rivers resulted in a coelacanth habitat disjunction that allowed populations on either side of India to diverge.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Indonesian coelacanth; Partitioned Bayesian analysis; Molecular clock; Basal Actinopterygii; Mitogenome

1. Introduction

Coelacanths were believed to have gone extinct more than 80 Mya until the sensational rediscovery of one

surviving member of this lineage, *Latimeria chalumnae*, in 1938 (Thomson, 1991). Since then, more than 200 coelacanths have been caught off the Comoro archipelago near the eastern coast of Africa in the Indian Ocean (Forey, 1998). During the period of September 1997 through July 1998, two coelacanths were captured off the coast of Manado, Sulawesi, Indonesia, some 10,000 km east of the southwestern Indian Ocean (Erdmann et al., 1998). These coelacanths are the first individuals recorded from a location outside the western Indian Ocean. The extensive interviews with Indonesian fishermen, combined with the vast distance from the Comoro archipelago, supported the idea that the Indonesian

Abbreviations: ATPase 6 and 8, ATPase subunits 6 and 8; bp, base pair(s); COI–III, cytochrome *c* oxidase subunits I–III; cyt *b*, cytochrome *b*; DMSO, dimethyl sulfoxide; Mya, million years ago; Myr, million years; NDI–6, 4L, NADH dehydrogenase subunits 1–6, 4L; PCR, polymerase chain reaction; tRNA, transfer RNA; 12S rRNA and 16S rRNA, 12S and 16S ribosomal RNA.

* Corresponding author. Tel.: +81 3 5351 6396; fax: +81 3 5351 6488.

E-mail address: jinoue@ori.u-tokyo.ac.jp (J.G. Inoue).

coelacanths are part of an established north Sulawesi population, and not simply waifs from the Comoran population (Forey, 1998).

Taxonomic confusion raised by overlapping morphological variations has posed problems in relation to coelacanth dispersal and biogeography. Pouyaud et al. (1999) described the Indonesian coelacanth as a new species, *Latimeria menadoensis*, based on nine morphological and meristic differences. However, Erdmann et al. (1999) suggested that the Indonesian coelacanth is morphologically similar to the Comoran coelacanth based on a preliminary comparison of external morphological measurements. Based on a survey of the literature, Holder et al. (1999) also stated that the case for morphological differentiation of *L. menadoensis* is much more tenuous than originally reported (Pouyaud et al., 1999) and concluded that one important morphological character that should be considered in future examination is that of scale ornamentation.

Two independent research groups have published results from divergence time estimation of the two coelacanths using partial mitochondrial gene sequences (6.3–4.7 Mya: Holder et al., 1999; 1.3 Mya: Pouyaud et al., 1999). Nevertheless, divergence time between the two coelacanths has remained ambiguous. In general, lineage-specific variation in rate of molecular evolution complicates molecular dating because a calibration rate estimated from one lineage may not be an accurate representation of the rate in other lineages (Bromham and Penny, 2003). These two studies, however, used rates of evolution from amphibians or teleosts for estimation without conducting the test of rate variation among lineages. Moreover, estimation of divergence time is generally more difficult than reconstruction of a phylogenetic tree, because no gene would evolve at a constant rate (Glazko and Nei, 2003). Considering the effect of rate variation among small number of genes, it is no wonder that the analyses based on partial mitochondrial gene sequences may be significantly biased.

In this study, we determined whole mitochondrial genome for Indonesian coelacanth *L. menadoensis* and compared the new sequence to that already reported for Comoran coelacanth *L. chalumnae*. Recent authors have used many genes to estimate divergence times in the hope of reducing the effect of rate variation (Nei and Glazko, 2002). We used whole mitochondrial genome sequences to estimate the divergence time between the two coelacanths because they have been demonstrated in recent studies as being useful for estimating the divergence times among basal lineages within tetrapods (Kumazawa et al., 2004) and within primates (Schrage and Russo, 2003). Two distinct methodologies were employed to estimate divergence time: the partitioned Bayesian approach and the molecular clock approach. Based on the molecular evidence, we evaluated several alternative hypotheses about the speciation of the two coelacanths.

2. Materials and methods

2.1. DNA extraction

Genomic DNA of the *L. menadoensis* was extracted from pieces of gills that were preserved in DMSO or ethanol by using the standard protocol.

2.2. PCR and sequencing

The mitochondrial genome of the *L. menadoensis* was amplified in the entirety using a long PCR technique. Four fish-versatile long PCR primers (S-LA-16S-L+H1065-12S and L12321-Leu+S-LA-16S-H; for locations of these primers, see Miya and Nishida, 2000) were used to amplify the entire mitochondrial genome in two reactions.

The long-PCR products were diluted with TE buffer (1:19) for subsequent use as PCR templates, except for a region intervening between the two long-PCR primers (between S-LA-16S-H and S-LA-16S-L), in which total genomic DNA was used alternatively. A total of 81 fish-versatile and five species-specific PCR primers (Supplement) was used in various combinations to amplify contiguous, overlapping segments of the entire mitochondrial genome for *L. menadoensis*. Five species-specific primers were designed in cases where no appropriate fish-versatile primers were available. Long PCR and subsequent short PCR were carried out as previously described (e.g. Miya and Nishida, 2000; Inoue et al., 2003).

Double-stranded PCR products, purified using a Pre-Sequencing Kit (USB), were subsequently used for direct cycle sequencing with dye-labeled terminators (Applied Biosystems). Primers used were the same as those for PCR. All sequencing reactions were performed according to the manufacturer's instructions. Labeled fragments were analyzed on a Model 3100 DNA sequencer (Applied Biosystems).

2.3. Sequence analysis

The DNA sequences were edited and analyzed with EditView ver. 1.0.1, AutoAssembler ver. 2.1 (Applied Biosystems), and DNASIS ver. 3.2 (Hitachi Software Engineering). Locations of the 13 protein-coding genes were determined by comparisons of DNA or amino acid sequences of bony fish mitochondrial genomes. The 22 tRNA genes were identified by their cloverleaf secondary structures and anticodon sequences. The two rRNA genes were identified by sequence similarity. Sequence data is available from DDBJ/EMBL/GenBank with accession number AP006858.

2.4. Alignment

The mitochondrial genome sequences from the Indonesian (this study) and Comoran (Zardoya and Meyer, 1997)

coelacanths were aligned with those of the 26 actinopterygians that fully represent major basal actinopterygian lineages and a shark used in Inoue et al. (2003). Protein-coding and tRNA gene sequences were aligned manually using MacClade ver. 4.06 (Maddison and Maddison, 2003). Amino acids were used for alignments of the protein-coding genes and secondary structure for alignment of tRNA genes. Since strictly secondary-structure based alignment for the two rRNA genes was impractical for the large data set, we alternatively employed machine alignment that would minimize erroneous assessment of positional homology of the rRNA molecules. The two rRNA gene (12S and 16S rRNA) sequences were aligned using ProAlign ver. 0.5 (Löytynoja and Milinkovitch, 2003) and those regions with posterior probabilities of $\geq 70\%$ were used in the analysis.

The ND6 gene was not used in the analysis because of its heterogeneous base composition (e.g. Kumazawa et al., 2004). Ambiguous alignment regions, such as the 5' and 3' ends of several protein-coding genes and loop regions of several tRNA genes, were excluded, leaving a total of 13,872 available nucleotide positions (10,635, 1434, and 1803 positions for protein-coding, tRNA, and rRNA genes, respectively) for the analysis. The “saturation” at 3rd codon

positions in the protein-coding genes was alleviated from the analyses by translating nucleotide sequences into amino acid sequences or simply excluding 3rd codon positions in the protein-coding genes. Two data sets were used in our analyses: data set #1: concatenated amino acid sequences from 12 protein-coding genes (total 3545 positions); data set #2: concatenated nucleotide sequences from 12 protein-coding genes (without 3rd codon positions), 22 transfer RNA genes, and two ribosomal RNA genes (total 10,327 positions).

2.5. Partitioned Bayesian analysis

The analysis of divergence time was conducted with two different approaches, the partitioned Bayesian approach (Thorne and Kishino, 2002) and the molecular clock approach (for a recent review, see Hedges and Kumar, 2003). A shark, *Scyliorhinus canicula*, was used as an outgroup. As a reference point for dating, the divergence time between sarcopterygians and actinopterygians (450 Mya) was used for the age of root node following previous analyses based on both fossils and molecules (see Kumazawa and Nishida, 2000; Hedges and Kumar, 2003).

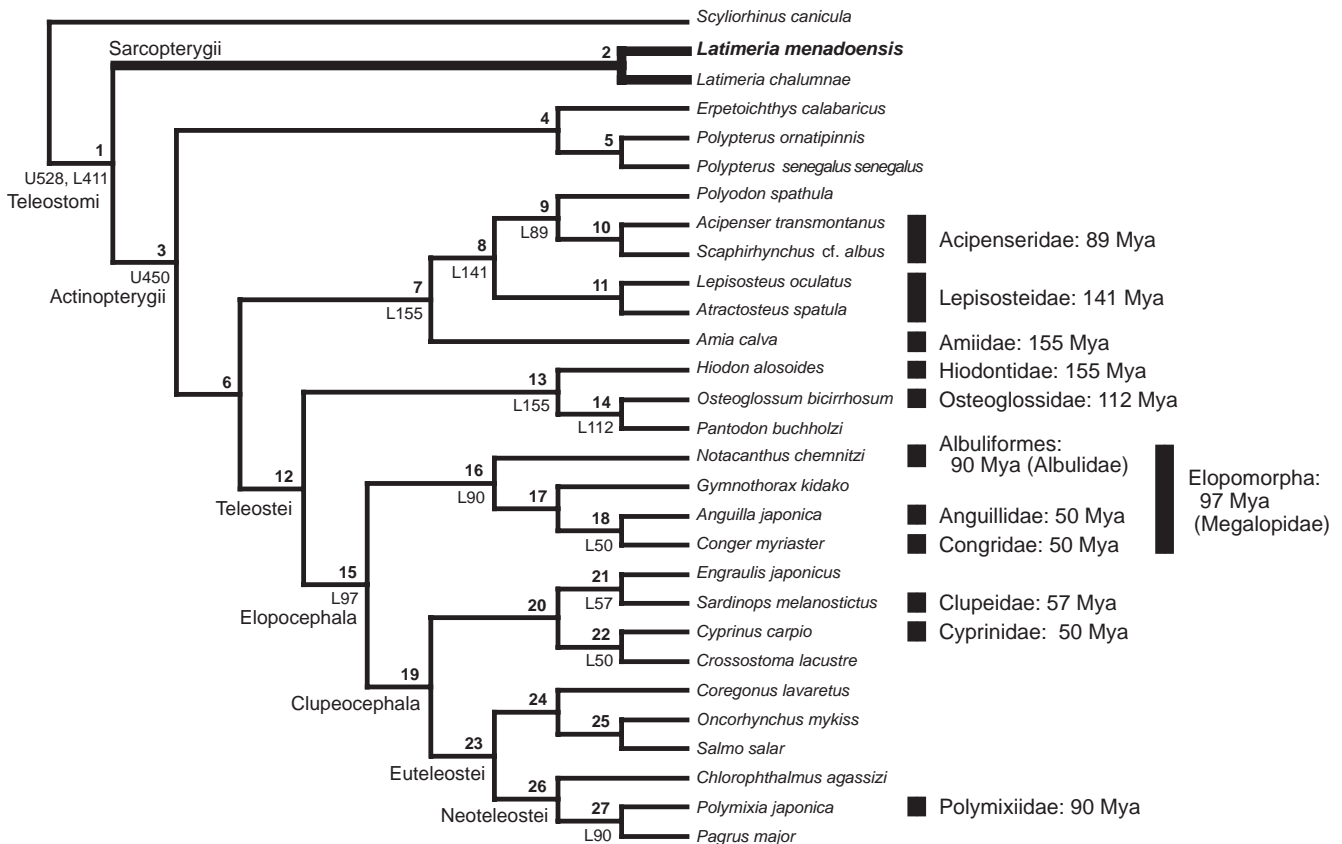


Fig. 1. Phylogenetic tree of the 28 bony fishes plus an outgroup and prior distributions of divergence times. We assumed that the tree topology is known and fixed, with reference to the recent mitochondrial genomic analysis of basal actinopterygian phylogeny by Inoue et al. (2003). Ages for earliest fossil records are shown to right of the tree. The divergence time between sarcopterygians and actinopterygians (Node 1), i.e. 450 Mya, was used for the age of root node following previous analyses based on both fossils and molecules (see Kumazawa and Nishida, 2000; Hedges and Kumar, 2003). In the partitioned Bayesian approach, we used the additional two maximum (U) and 12 minimum (L) constraints according to the fossil and molecular information (Table 1).

The program Thornian Time Traveller ver. 1.0 (T3; see <http://abacus.gene.ucl.ac.uk/>) was used to estimate the divergence time following the partitioned Bayesian method of Thorne and Kishino (2002). We set 12 (data set #1: 12 protein-coding genes) and 4 (data set #2: 1st, 2nd, codon positions, tRNA, and rRNA genes) partitions depending on the data sets. For data set #2, we assumed that functional constraints on sequence evolution are more similar within codon positions (or types of molecules) across genes than across codon positions (or types of molecules) within genes. Branch lengths were estimated with the *estbNewAA* and *estbNew* programs of T3 for the data sets #1 and #2, respectively, in conjunction with the tree topology shown in Fig. 1. We used the *mtmam* (Yang et al., 1998) model for data set #1 and the F84 (Felsenstein, 1984) model incorporating among-site rate variation modeled by a gamma distribution (Yang, 1994) of sequence evolution (F84+ Γ model) for data set #2. For data set #2, we used the *baseml* program in the PAML ver. 3.14 package (Yang, 1997) to obtain estimates of the transition/transversion rate ratio and the rates for site classes under the discrete-gamma model of rates among sites under the F84+ Γ model.

Divergence time was estimated using the program *multidivtime* of T3. In each case, Markov chain Monte Carlo (MCMC) approximations were obtained with a burnin period of 100,000 proposal cycles. Thereafter, samples of the Markov chain were taken every 100 cycles until a total of 10,000 samples were obtained. To diagnose possible failure of the Markov chains to converge to their stationary distribution, we performed at least two replicate MCMC runs with different initial starting points for each analysis. Application of the *multidivtime* program requires a value for the mean of the prior distribution for the time separating the ingroup root from the present (*rttm*). We used a conservative estimate of 450 Mya for *rttm* (see above). Other settings for running *multidivtime* were as follows: standard deviation of *rttm* (*rttmsd*)=0.5×*rttm*; *rrate*=*X*/*rttm*, where *rrate* is the mean of prior distribution for the rate at the root node and *X* is the median amount of evolution from the ingroup root to the ingroup tips; *rratesd*=0.5×*rrate*, where *rratesd* is the standard deviation of *rrate*; *rttm*×*brownmean*=1, where *brownmean* is the mean of the prior distribution for the autocorrelation parameter (*v*); and *brownsd*=*brownmean*, where *brownsd* is the standard deviation of the prior distribution for *v*.

The *multidivtime* program allows for both minimum and maximum fossil constraints. Whereas minima are often based on earliest occurrences in the fossil record, maxima are intrinsically more difficult to estimate. With the exception of Nodes 1 and 3 (Table 1, Fig. 1), we could not use the maximum because fossils of teleosts are not considered well preserved in general (Benton, 1993; Kumazawa and Nishida, 2000). Accordingly, we used the two maximum and 12 minimum constraints for the divergence time estimation (Table 1, Fig. 1).

2.6. Molecular clock approach

In the molecular clock analysis, we estimated the divergence time of the two coelacanths for data set #1, because DNA sequences are usually less conserved than protein sequences and the substitution pattern in DNA sequences varies extensively with codon position (Glazko and Nei, 2003). In order to use molecular sequence data as the molecular clock, the homogeneity of modes and rates of sequence evolution among lineages must be carefully examined. The two cluster test (LINTREE; Takezaki et al., 1995) was then conducted using the gamma-corrected amino-Poisson (Poisson+ Γ) distances among the 28 taxa. Since a shark was used as an outgroup, evolutionary rates were compared between clusters created by all bony fish nodes of the tree topology shown in Fig. 1. Faster or more slowly evolving sequences were removed and the procedure was repeated until the data set contained only sequences evolving at a similar rate. After such processes, we assumed that the tree topology is known and fixed, with reference to the recent mitochondrial genome analysis of the basal actinopterygian phylogeny by Inoue et al. (2003). Estimation of divergence time was based on *mtREV* (Adachi and Hasegawa, 1996)+ Γ distance of data set #1. The *mtREV*+ Γ distances were obtained with *TREE-PUZZLE* ver. 5.0

Table 1

Maximum (U) and minimum (L) constrains (Mya) and calibration information for nodes in Fig. 1

| Node | Constraints | Calibration information |
|------|-------------|---|
| 1 | U528 | The estimated divergence time between chondrichthyans and osteichthyans (528 Mya) based on both fossils and molecules (Kumazawa and Nishida, 2000; Hedges and Kumar, 2003) |
| 1 | L411 | The <i>Andreolepis</i> fossil (Actinopterygii) from Ludfordian (Silurian) ^a |
| 3 | U450 | The estimated divergence time between sarcopterygians and actinopterygians (450 Mya) based on both fossils and molecules (Kumazawa and Nishida, 2000; Hedges and Kumar, 2003) |
| 7 | L155 | The amiid fossil from Oxfordian (Jurassic) ^a |
| 8 | L141 | The lepisosteoid fossil from Berriasian (Cretaceous) ^a |
| 9 | L89 | The acipenserid fossil from Turonian (Cretaceous) ^a |
| 13 | L155 | The hiodontid fossil from the Oxfordian (Jurassic) ^a |
| 14 | L112 | The osteoglossoid fossil from the Aptian (Cretaceous) ^a |
| 15 | L97 | The megalopid fossil (Elopomorpha) from the Albian (Cretaceous) ^a |
| 16 | L90 | The albuloid fossil from the Cenomanian (Cretaceous) ^a |
| 18 | L50 | The anguillid and congrid fossils from the Ypresian (Tertiary) ^a |
| 21 | L57 | The clupeid fossil from the Thanetian (Tertiary) ^a |
| 22 | L50 | The cyprinid fossil from the Ypresian (Tertiary) ^a |
| 27 | L90 | The polymixiid fossil from the Cenomanian (Cretaceous) ^a |

^a Minima (L) are based on earliest occurrences in the fossil records (Benton, 1993).

(Schmidt et al., 2002), using a gamma parameter ($\alpha=0.33$) estimated from the data set.

3. Results

3.1. Genome organization

The complete L-strand nucleotide sequence of *L. menadoensis* has been registered in DDBJ/EMBL/GenBank under the accession number AP006858. The total length of the *L. menadoensis* mitochondrial genome was 16,446 bp. The genome contents included 13 protein-coding, 22 tRNA, and two rRNA genes, and a control region, as found in other vertebrates. Also, as in other vertebrates, most genes of this species were encoded on the H-strand, except for the ND6 and eight tRNA genes, all genes being similar in length to those in other bony fishes. The gene order was identical to those so far obtained in other typical vertebrates. The sequence was identical to that recently reported for six genes (cyt *b*, tRNA^{Pro}, tRNA^{Thr}, tRNA^{Val}, 12S rRNA, and 16S rRNA) and the control region of mitochondrial genome from *L. menadoensis* (Holder et al., 1999).

Nucleotide or amino acid sequences for the homologous genes of the mitochondrial genomes were compared between *L. menadoensis* and *L. chalumnae* (Table 2). Lower values of nucleotide sequence differences between *L. menadoensis* and *L. chalumnae* were observed in 12S rRNA (2.35%), 16S rRNA (3.19%), and tRNA (3.23%) genes. In general, the ranges of nucleotide differences were much higher in genes, which encode NADH dehydrogenase subunits (4.63–6.02%) with the exception of ND4 gene (3.77%). The three most variable mitochondrial genes were ND1 (5.56%), ND2 (5.92%), and ND3 (6.02%). The

Table 3

Divergence time estimates of bony fish lineages based on the partitioned Bayesian approach for data set #1 of nodes in Fig. 2A and for data set #2 of the corresponding node (Mya)

| Node | Data set #1 | Data set #2 |
|------|----------------------|----------------------|
| 1 | 438.0 (412.0, 480.3) | 450.5 (413.2, 494.9) |
| 2 | 34.4 (25.7, 44.2) | 30.0 (24.0, 37.2) |
| 3 | 407.0 (376.2, 445.8) | 414.8 (374.3, 448.4) |
| 4 | 81.9 (66.2, 100.6) | 68.3 (54.0, 85.0) |
| 5 | 50.2 (38.4, 64.3) | 43.1 (32.2, 56.4) |
| 6 | 342.5 (310.0, 380.8) | 376.4 (336.7, 413.4) |
| 7 | 327.4 (294.9, 365.5) | 359.8 (319.7, 397.8) |
| 8 | 312.1 (279.4, 350.6) | 346.9 (306.8, 385.1) |
| 9 | 114.1 (91.8, 143.7) | 145.2 (108.1, 185.9) |
| 10 | 65.3 (46.8, 88.7) | 79.0 (53.7, 110.4) |
| 11 | 33.1 (23.9, 44.4) | 66.2 (49.2, 87.8) |
| 12 | 284.7 (253.2, 320.4) | 333.8 (294.9, 371.6) |
| 13 | 274.5 (243.9, 309.2) | 280.8 (243.3, 317.7) |
| 14 | 200.1 (170.6, 233.4) | 192.6 (159.3, 227.8) |
| 15 | 265.0 (233.7, 300.0) | 314.7 (275.8, 351.8) |
| 16 | 228.3 (198.5, 261.9) | 263.8 (227.3, 301.1) |
| 17 | 166.2 (139.9, 195.6) | 202.6 (169.5, 237.3) |
| 18 | 149.1 (124.2, 177.3) | 175.8 (145.2, 207.7) |
| 19 | 230.4 (200.3, 263.6) | 277.7 (240.9, 313.9) |
| 20 | 200.8 (172.3, 232.5) | 239.0 (203.9, 274.8) |
| 21 | 118.0 (96.1, 142.7) | 140.5 (112.4, 171.1) |
| 22 | 133.4 (110.2, 159.7) | 154.7 (123.2, 187.4) |
| 23 | 190.7 (163.7, 221.1) | 231.6 (197.1, 267.0) |
| 24 | 78.8 (60.6, 99.9) | 111.6 (84.8, 140.9) |
| 25 | 48.0 (33.7, 64.9) | 77.0 (55.6, 101.8) |
| 26 | 160.5 (135.1, 189.2) | 193.7 (161.4, 228.4) |
| 27 | 131.2 (108.1, 156.8) | 159.3 (129.9, 191.0) |

Numbers in the parentheses indicate 95% credibility intervals of divergence times.

nucleotide difference of control regions was 6.27%. In the protein coding genes, variations at 3rd codon positions (4.78–13.79%) were higher than those at 1st (1.36–5.36%) and 2nd (0–2.01%) codon positions as expected. Amino

Table 2

Uncorrected nucleotide and amino acid sequence differences (%) for the homologous regions of the mitochondrial genomes between *Latimeria menadoensis* and *Latimeria chalumnae*

| Features | Nucleotide | | | | Amino acid |
|----------------|------------|-----------|-----------|-----------|------------|
| | 1st | 2nd | 3rd | Total | |
| ATP6 | 4.55 | 1.83 | 7.32 | 4.53 | 5.46 |
| ATP8 | 5.36 | 0.00 | 7.14 | 4.17 | 1.79 |
| COI | 1.36 | 0.19 | 10.30 | 3.95 | 0.78 |
| COII | 1.73 | 0.87 | 4.78 | 2.46 | 1.74 |
| COIII | 3.44 | 1.91 | 6.87 | 4.07 | 4.22 |
| cyt <i>b</i> | 1.58 | 1.58 | 10.53 | 4.55 | 2.37 |
| ND1 | 4.01 | 0.62 | 12.04 | 5.56 | 3.40 |
| ND2 | 4.30 | 2.01 | 11.46 | 5.92 | 5.73 |
| ND3 | 2.56 | 1.72 | 13.79 | 6.02 | 2.59 |
| ND4 | 3.47 | 0.87 | 6.96 | 3.77 | 3.26 |
| ND4L | 3.03 | 0.00 | 12.12 | 5.05 | 3.03 |
| ND5 | 3.43 | 1.14 | 9.31 | 4.63 | 3.76 |
| ND6 | 5.17 | 1.15 | 8.62 | 4.98 | 5.75 |
| 12S rRNA | – | – | – | 2.35 | – |
| 16S rRNA | – | – | – | 3.19 | – |
| 22 tRNA | – | – | – | 3.23 | – |
| Control region | – | – | – | 6.27 | – |
| Mean±S.D. | 3.38±1.31 | 1.07±0.72 | 9.33±2.64 | 4.39±1.18 | 3.38±1.58 |

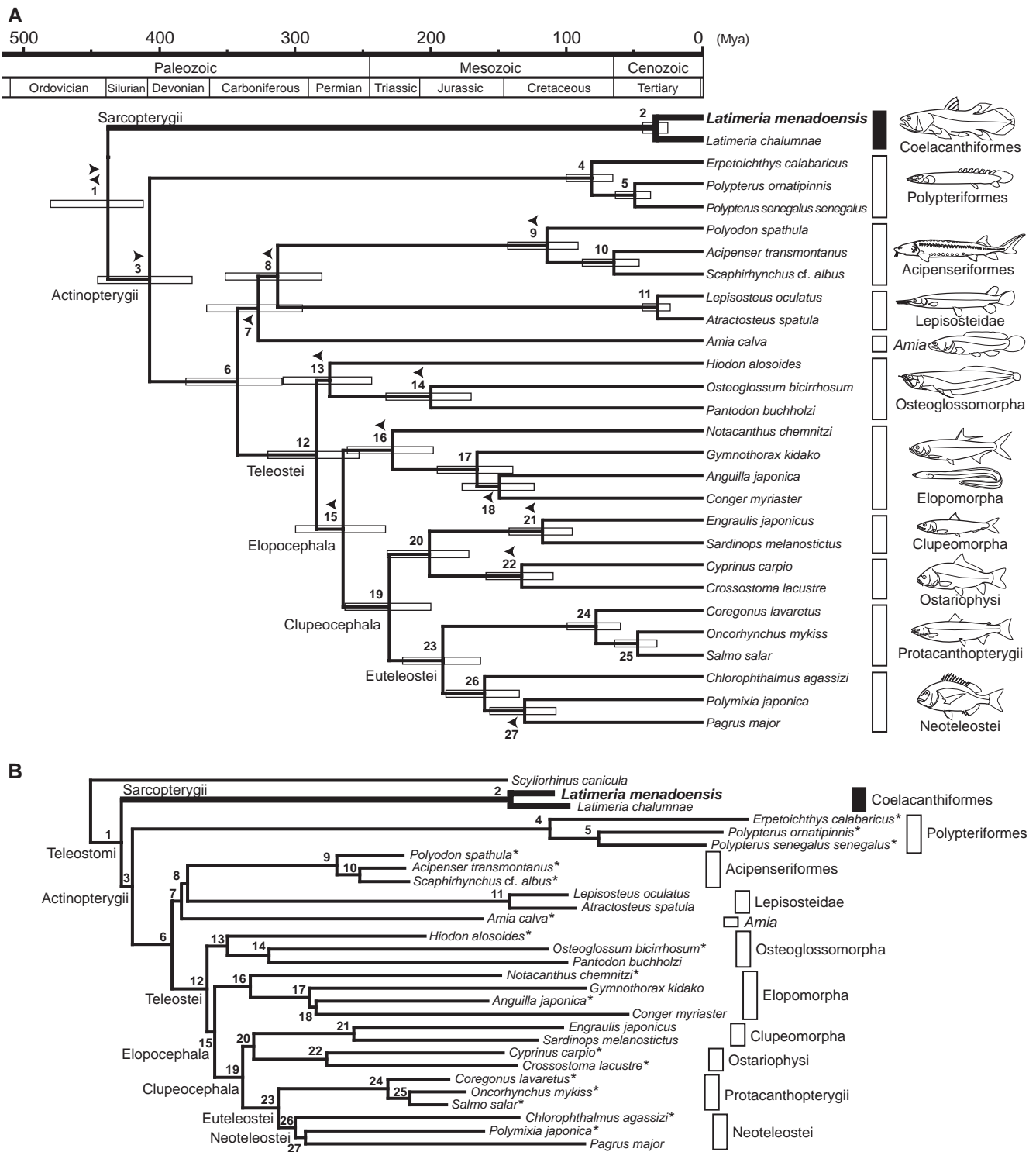


Fig. 2. (A) Posterior distributions of the divergence times among basal bony fishes based on the partitioned Bayesian approach (Thorne and Kishino, 2002) for data set #1 [concatenated amino acid sequences from 12 protein-coding genes (total 3545 positions)]. A shark, *Scyliorhinus canicula*, was used as an outgroup. The program Thornian Time Traveller ver. 1.0 (T3; see <http://abacus.gene.ucl.ac.uk/>) was used to estimate divergence times in conjunction with the tree topology shown in Fig. 1 using the mtmam model of sequence evolution. The horizontal rectangles represent the estimated 95% credibility intervals of divergence times. Arrowheads indicate the nodes under maximum or minimum constraints (see Table 1, Fig. 1). (B) Neighbor-joining tree constructed from data set #1 using a shark as an outgroup. The Poisson+ Γ distance was used with njboot in LINTREE package (Takezaki et al., 1995). With the exception of two coelacanths, the tree topology was congruent with those found by Inoue et al. (2003). The two-cluster test showed, with a 99% significance level, that the substitution rates for the 18 actinopterygian lineages (asterisks) are significantly different from those for the two coelacanths.

acid sequence differences of protein-coding genes ranged from 0.78% (COI gene) to 5.75% (ND6 gene).

3.2. Divergence time estimation

The partitioned Bayesian approach estimated the divergence time between the two coelacanths at 34.4 Mya with the 95% credibility interval of 25.7–44.2 Mya for data set #1 and 30.0 Mya with the 95% credibility interval of 24.0–37.2 Mya for data set #2 (Node 2 in Table 3 and Fig. 2A).

The two-cluster test was conducted using the Poisson+ Γ distances among the 28 taxa for data set #1. The test showed, with a 99% significance level, that the substitution rates for the 18 actinopterygian lineages were significantly different from those for the two coelacanths (Fig. 2B). We thus excluded the 18 actinopterygians from the analysis. Using the sarcopterygian–actinopterygian split as a calibration point (450 Mya), a time estimate of 39.1 Mya for the split of the two coelacanths was obtained for data set #1 with mtREV+ Γ model.

The molecular clock of bony fishes for amino acid sequences of the mitochondrial ND2 and *cyt b* genes was calibrated by Kumazawa and Nishida (2000) using three calibration points, the divergence of African and Neotropical cichlids at the time of the continental breakup of the African and South American landmasses at 100 Mya, sarcopterygian versus actinopterygian divergence at 450 Mya, and chondrichthyan versus osteichthyan divergence at 528 Mya. The amino acid sequences of the ND2 and *cyt b* genes from Indonesian coelacanth were added to the data matrix of Kumazawa and Nishida (2000), which includes Comoran coelacanth, 23 actinopterygians, and five outgroups. Two-cluster test showed that the rates for the two coelacanths were not significantly different from those for the actinopterygians with a 99% significance level (with the exception of that for the northern barramundi). When the mtREV+ Γ distance was used with the molecular clock (5.4×10^{-4} substitutions/site/My) of Kumazawa and Nishida (2000), we obtained 36.8 Mya for the divergence between the two coelacanths.

4. Discussion

4.1. Divergence time estimation

Divergence time estimation between *L. menadoensis* and *L. chalumnae* fell in the range of 40–30 Mya. Although the level of nucleotide sequence divergences for the homologous genes between the two coelacanths (Table 2) are similar to those reported in Holder et al. (1999) and Pouyaud et al. (1999), our estimate of the divergence time between the two coelacanths is much older than those of the previous studies. Pouyaud et al. (1999) reported that the estimated time of divergence was around 1.3 Mya based on *cyt b* and 12S rRNA gene sequences (1829 bp) assuming

that the *cyt b* gene was evolving at a substitution rate of 2.0×10^{-2} and 12S rRNA gene at 1.0×10^{-2} (all rates in substitutions/site/My). Holder et al. (1999) calculated the date to be 6.3–4.7 Mya based on the nucleotide sequences from the *cyt b* and two rRNA genes and the control region (4527 bp) using a substitution rate of 7.7×10^{-3} for *cyt b* gene, 3.8×10^{-3} for rRNA genes (Caccone et al., 1997), and 2.0×10^{-2} for control region (Brown et al., 1979) (all rates in substitutions/site/My). These two studies used rates of evolution from amphibians or teleosts for estimation without conducting the test of rate variation among lineages.

In order to address whether or not nucleotide sequences from *cyt b* or rRNA genes alone were suitable for dating the split between the two coelacanths, we reestimated the divergence time employing the partitioned Bayesian approach for unambiguously aligned nucleotide sequences of 1st and 2nd codon positions of the *cyt b* gene (754 bp) and the two rRNA genes (1803 bp) from the 29 species (Fig. 1), using the sarcopterygian–actinopterygian split as a calibration point (450 Mya) with two maximum and 12 minimum constraints (Table 1). We set two (1st and 2nd codon positions of *cyt b* gene) and one (rRNA genes) partitions depending on the data set used. The partitioned Bayesian approach estimated the data at 27.8 Mya with the 95% credibility interval of 6.3–53.4 Mya based on *cyt b* gene sequences and 20.1 Mya with the 95% credibility interval of 11.2–31.8 Mya based on rRNA gene sequences. Although, on the basis of nucleotide sequences of *cyt b* or two rRNA genes, we obtained data estimates for the split between the two coelacanths closer to the date based on the mitochondrial genome data (about 40–30 Mya), the credibility intervals were more narrow for the multigene analyses than for the single-gene analyses as expected.

The partitioned Bayesian approach with the whole mitochondrial genome data also estimated the times of major divergence events within the actinopterygians (data set #1/#2): Lepisosteidae–Teleostei (Node 6 in Table 3 and Fig. 2A) at 343/376 Mya; Osteoglossomorpha–Elopocephala (Node 12) at 285/334 Mya. These estimates agree with previous molecular studies (*Amia*–Teleostei at 404 Mya: Kumazawa et al., 1999; Osteoglossomorpha–Elopocephala at 341 Mya: Kumazawa and Nishida, 2000). However, fossil records for the teleosts and the related extinct groups have not been found before the Late Triassic (235 Mya) (Benton, 1993). Thus, there appears to be a time gap (ca. 290–235 Mya) between the molecular and fossil evidence. We interpret this apparent discrepancy to be indicative of paucity of teleostean fossil records rather than the inferiority of our molecular time estimates.

The estimated divergence time (about 40–30 Mya) between the two coelacanths is comparable to those between some major lineages of mammals, for example, toothed versus baleen whales (27 Mya: Hasegawa et al., 2003; 30 Mya: Springer et al., 2003) or humans versus New World monkeys (33 Mya: Glazko and Nei, 2003; 35 Mya: Schrago and Russo, 2003). The split between the two coelacanths

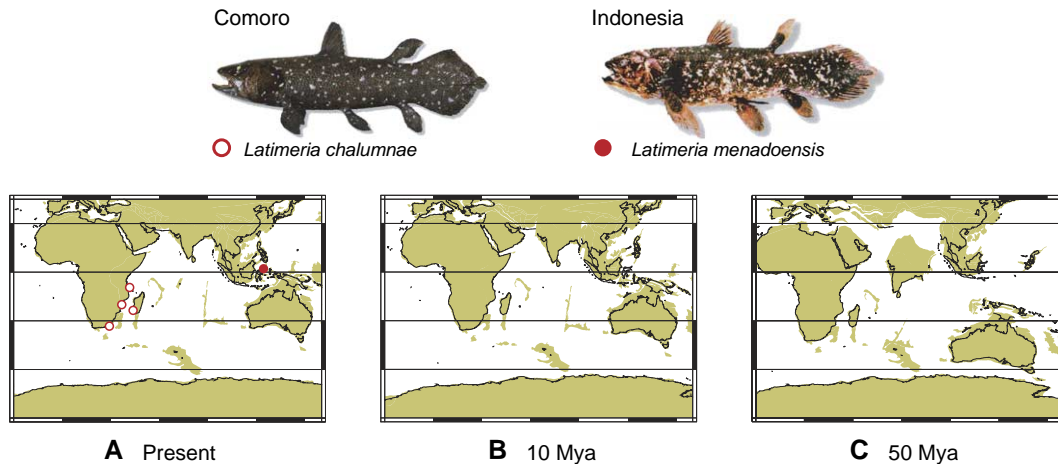


Fig. 3. Geographical maps of Indian Ocean at (A) present, (B) 10, and (C) 50 Mya based on the Plate Tectonic Reconstruction Service of the Ocean Drilling Stratigraphic Network (Hay et al., 1999). The present-day shorelines are shown in black and the plate fragments are shown in ochre. Gondwana started to break up in the Jurassic (165–150 Mya), when India began to drift away from Australia–East Antarctica. India separated from Madagascar in the late Cretaceous (88–84 Mya) and began drifting northward, eventually to collide with Asia at about 50 Mya (C) (Sanmartin and Ronquist, 2004). On the other hand, complete separation between Australia and Antarctica occurred in the early Oligocene (35 Mya), and Australia began to drift rapidly toward Asia. The collision of the Australian and Asian plate in the late Miocene–Pliocene (10 Mya) resulted in the formation of the Indo–Australian Arc (B). The thumbnail photographs of coelacanths are reproduced through the courtesy of the Shogakukan.

occurred during the early Oligocene, which was much deeper than expected from the fact that they were classified within the same genus *Latimeria*. In contrast, the divergence between suborders of cetaceans or infraorders of primates is reasonably deep in light of their classificatory status. These observations highlight the exceptionally high morphological conservation among the *Latimeria* species.

4.2. Evolutionary history of the two coelacanths

The distribution pattern of two living coelacanths raises an important question about their dispersal and biogeography (Fig. 3). Assuming that the most recent ancestor of the two living coelacanths was distributed along the coasts of Africa through Eurasia (Springer, 1999), three alternative models have been proposed to explain the evolutionary history of the two coelacanths. First model postulates the possibility that there is gene flow between western Indian Ocean and Indonesian populations of *Latimeria* (Fig. 3A). Second model proposes that the massive tectonic changes that led to the formation of the Indo–Australian Arc (10 Mya) and resultant separation of the Indian Ocean and Pacific Ocean (Fig. 3B) caused a barrier to gene flow between populations of the two coelacanths during the Miocene (Holder et al., 1999). However, the observed level of differentiation between the mitochondrial gene sequences of the two coelacanths (Holder et al., 1999; Pouyaud et al., 1999; this study) implies a long history of genetic separation and the formation of the Indo–Australian Arc was initiated significantly later (late Miocene through the Pliocene, 10–2 Mya) than our molecular estimate (about 40–30 Mya).

Third model postulates that the collision of India with Eurasia (50 Mya) and the subsequent siltation caused by the formation of major rivers formed a barrier isolating the

distributions of coelacanths (Fig. 3C) (Springer, 1999). Springer (1999) also suggested that the isolation of the Comoran and Indonesian coelacanth populations probably began several My after the docking of India with Eurasia because the formation of the Himalayas and associated major river systems would have taken considerable time to develop. The estimated divergence time between the two coelacanths (about 40–30 Mya) makes Springer's biogeographic hypothesis as the most likely hypothesis.

Acknowledgements

We thank Y. Kumazawa, M. Yamaguchi, K. Mabuchi, Y. N. Yamanoue, and M.M. Yamauchi for providing useful information and discussions on the results reported herein and M.K. Moosa, Susetiono, H. Ida, H. Senou, Kanagawa Prefectural Museum of Natural History, and Shogakukan Inc. for permission to use photographs. This study was supported in part by Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists (07304), Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology, Japan (12NP0201, 13556028, 13640711 15380131, and 15570090), and “Research for the Future” Program No. JSPS-RFTF 97L00901 from the Japan Society for the Promotion of Science.

Appendix A

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2005.01.008.

References

- Adachi, J., Hasegawa, M., 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* 42, 459–468.
- Benton, M.J., 1993. *The Fossil Record*, vol. 2. Chapman & Hall, London.
- Bromham, L., Penny, D., 2003. The modern molecular clock. *Nat. Rev. Gen.* 4, 216–224.
- Brown, W.M., George, Jr, M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 76, 1967–1971.
- Caccone, A., Milinkovitch, M.C., Sbordoni, V., Powell, J.R., 1997. Mitochondrial DNA rates and biogeography in European newts (Genus *Euproctus*). *Syst. Biol.* 46, 126–144.
- Erdmann, M.V., Caldwell, R.L., Moosa, M.K., 1998. Indonesian ‘king of the sea’ discovered. *Nature* 395, 335.
- Erdmann, M.V., Caldwell, R.L., Jewett, S.L., Tjakrawidjaja, A., 1999. The second recorded living coelacanth from north Sulawesi. *Environ. Biol. Fishes* 54, 445–451.
- Felsenstein, J., 1984. Distance methods for inferring phylogenies: a justification. *Evolution* 38, 16–24.
- Forey, P.L., 1998. A home from home for coelacanths. *Nature* 395, 319–320.
- Glazko, G.V., Nei, M., 2003. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* 20, 424–434.
- Hasegawa, M., Thorne, J.L., Kishino, H., 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* 78, 267–283.
- Hay, W.W., et al., 1999. Alternative global cretaceous paleogeography. In: Barrera, E., Johnson, C. (Eds.), *The evolution of cretaceous ocean/climate systems*, Geological Society of America Special Paper, vol. 332. Boulder, pp. 1–47.
- Hedges, S.B., Kumar, S., 2003. Genomic clocks and evolutionary timescales. *Trends Genet.* 19, 200–206.
- Holder, M.T., Erdmann, M.V., Wilcox, T.P., Caldwell, R.L., Hillis, D.M., 1999. Two living species of coelacanths? *Proc. Natl. Acad. Sci. U. S. A.* 96, 12616–12620.
- Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the “ancient fish.” *Mol. Phylogenet. Evol.* 26, 110–120.
- Kumazawa, Y., Nishida, M., 2000. Molecular phylogeny of osteoglossoids: a new model for gondwanian origin and plate tectonic transportation of the Asian arowana. *Mol. Biol. Evol.* 17, 1869–1878.
- Kumazawa, Y., Yamaguchi, M., Nishida, M., 1999. Mitochondrial molecular clocks and the origin of euteleostean biodiversity: familial radiation of perciforms may have predated the Cretaceous/Tertiary boundary. In: Kato, M. (Ed.), *The Biology of Biodiversity*. Springer-Verlag, Tokyo, pp. 35–52.
- Kumazawa, Y., Azuma, Y., Nishida, M., 2004. Tempo of mitochondrial gene evolution: can mitochondrial DNA be used to date old divergences? *Endocytobiosis Cell Res.* 15, 136–142.
- Löytynoja, A., Milinkovitch, M.C., 2003. A hidden Markov model for progressive multiple alignment. *Bioinformatics* 19, 1505–1513.
- Maddison, D.R., Maddison, W.P., 2003. *MacClade Ver. 4.05*. Sinauer, Sunderland, MA.
- Miya, M., Nishida, M., 2000. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. *Mol. Phylogenet. Evol.* 17, 437–455.
- Nei, M., Glazko, G.V., 2002. Estimation of divergence times for a few mammalian and several primate species. *J. Heredity* 93, 157–164.
- Pouyaud, L., Wirjoatmodjo, S., Rachmatika, I., Tjakrawidjaja, A., Hadiaty, R., Hadie, W., 1999. A new species of coelacanth. *C. R. Acad. Sci. III* 322, 261–267.
- Sanmartin, I., Ronquist, F., 2004. Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Syst. Biol.* 53, 216–243.
- Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. Tree-puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504.
- Schrägo, C.G., Russo, C.A.M., 2003. Timing the origin of new world monkeys. *Mol. Biol. Evol.* 20, 1620–1625.
- Springer, V.G., 1999. Are the Indonesian and western Indian Ocean coelacanths conspecific: a prediction. *Environ. Biol. Fishes* 54, 453–456.
- Springer, M.S., Murphy, W.J., Eizirik, E., O’Brien, S.J., 2003. Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1056–1061.
- Takezaki, N., Rzhetsky, A., Nei, M., 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* 12, 823–833.
- Thomson, K.S., 1991. *Living Fossil: The Story of the Coelacanth*. W. W. Norton, New York.
- Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51, 689–702.
- Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39, 306–314.
- Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13, 555–556.
- Yang, Z., Nielsen, R., Hasegawa, M., 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.* 15, 1600–1611.
- Zardoya, R., Meyer, A., 1997. The complete DNA sequence of the mitochondrial genome of a “living fossil,” the coelacanth (*Latimeria chalumnae*). *Genetics* 146, 995–1010.