J Mol Evol (1994) 38:602–609

Journal of Molecular Evolution

Relationship Among Coelacanths, Lungfishes, and Tetrapods: A Phylogenetic Analysis Based on Mitochondrial Cytochrome Oxidase I Gene Sequences

Shin-ichi Yokobori,[1,3] Masami Hasegawa,[2] Takuya Ueda,[3] Norihiro Okada,[1] Kazuya Nishikawa,[1] Kimitsuna Watanabe[3]

[1] Department of Biological Sciences, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 227, Japan
[2] The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan
[3] Department of Industrial Chemistry, Faculty of Engineering, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Received: 15 July 1993 / Revised: 7 December 1993

Abstract. To clarify the relationship among coelacanths, lungfishes, and tetrapods, the amino acid sequences deduced from the nucleotide sequences of mitochondrial cytochrome oxidase subunit I (COI) genes were compared. The phylogenetic tree of these animals, including the coelacanth *Latimeria chalumnae* and the lungfish *Lepidosiren paradoxa,* was inferred by several methods. These analyses consistently indicate a coelacanth/lungfish clade, to which little attention has been paid by previous authors with the exception of some morphologists. Overall evidence of other mitochondrial genes reported previously and the results of this study equally support the coelacanth/lungfish and lungfish/tetrapod clades, ruling out the coelacanth/tetrapod clade.

Key words: Origin of tetrapods — Coelacanth *Latimeria chalumnae* — Lungfish *Lepidosiren paradoxa* — Cytochrome oxidase subunit I (COI) — Maximum likelihood inference of protein phylogeny

Introduction

The appearance of tetrapods was one of the most important events in vertebrate evolution. During landing, ancestral tetrapods had to acquire four legs to move against gravity and develop lungs for the respiration of air. Sarcopterygians (lobe-finned fishes) have been considered to be closely related to the ancestor of tetrapods (Romer 1966) because they have bony fins, and some genera of them have lungs for respiration. Sarcopterygians comprise three groups: extinct rhipidistians; coelacanths, including the "living fossil" *Latimeria chalumnae;* and lungfishes (dipnoans), including several living taxa such as *Lepidosiren paradoxa* (Forey 1988). As the extinct rhipidistians are not thought to be a monophyletic group (Forey 1988), the relationship among sarcopterygians (rhipidistians, coelacanths, and lungfishes) and tetrapods remains very problematical.

Three possible hypotheses concerning the relationship among extant groups (coelacanth, lungfishes, and tetrapods) have been proposed (Forey 1988) that respectively regard lungfishes and tetrapods (Miles 1977; Rosen et al. 1981; Forey 1987) (Fig. 1, Tree-1), coelacanths and tetrapods (Miles 1975; Hennig 1983; Fritsch 1987; Schultze 1987) (Fig. 1, Tree-2), and lungfishes and coelacanths (Northcutt 1987; Chang 1991) (Fig. 1, Tree-3) as sister groups. In this paper, we analyze the relationship of the three extant groups of coelacanth, lungfishes, and tetrapods, premising that the sarcopterygians and tetrapods form a clade from which ray-finned fishes are excluded as an outgroup.

Molecular data have been widely applied in deducing the phylogeny of living organisms. Recently, the nu-

Correspondence to: K. Watanabe

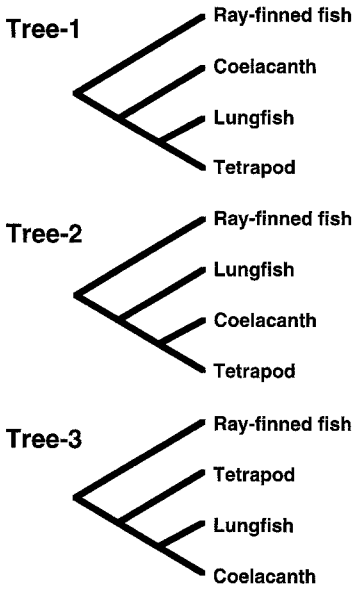

Fig. 1. Possible models exhibiting the relationship among the ray-finned fishes, coelacanths, lungfishes, and tetrapods. Tree-1 is constructed based on the phylogenetic model [ray-finned fish (coelacanth (lungfish, tetrapod))] (Miles 1977; Rosen et al. 1981; Forey 1987; Meyer and Wilson 1990; Meyer and Dolven 1992). Tree-2 is based on the model [ray-finned fish (lungfish (coelacanth, tetrapod))] (Miles 1975; Hennig 1983; Fritsch 1987; Schultze 1987; Gorr et al. 1991; see text). Tree-3 is based on the model [ray-finned fish ((lungfish, coelacanth) tetrapod)] (Northcutt 1987; Chang 1991).

cleotide and amino acid sequences of various genes were analyzed in an attempt to clarify the origin of tetrapods; mitochondrial 12S rRNA (Meyer and Wilson 1990; Meyer and Dolven 1992), mitochondrial apocytochrome b (cyt b) (Meyer and Wilson 1990; Nomark et al. 1991), hemoglobins (Gorr et al. 1991; and see also Forey 1991; Stock and Swofford 1991; Sharp et al. 1991; Meyer and Wilson 1991), and 18S rRNA (Stock et al. 1991) have been analyzed. However, the relationship among coelacanths, lungfishes, and tetrapods has remained obscure. Although a lungfish/tetrapod clade is suggested from the 12S rRNA data (Meyer and Wilson 1990; Meyer and Dolven 1992) as well as from the cyt b data (Meyer and Wilson 1990), other molecular data have been equivocal. Therefore, more such data are needed to clarify the origin of tetrapods.

In this study, we analyzed the mitochondrial cytochrome oxidase subunit I (COI) genes of the coelacanth *L. chalumnae* and the lungfish *L. paradoxa* to clarify the relationship among ray-finned fishes, lungfishes, coelacanth, and tetrapods. Previously, mitochondrial DNA data have been used mainly to analyze the phylogeny of closely related organisms, such as those within Hominoidea, because of the rapid evolutionary rate in mitochondrial DNA (Brown et al. 1982). However, it has been demonstrated that the nucleotide and amino acid sequences of the conservative mitochondrial genes are useful for analyzing relationships among distantly related organisms (Kocher et al. 1989; Miyamoto and Boyle 1989; Irwin et al. 1990; Kumazawa and Nishida 1993). Both the nucleotide and amino acid sequences of COI are very conservative among metazoa (Jacobs et al. 1988a), and hence this gene is expected to be useful for our purpose. One of the advantages of using mitochondrial genes in phylogenetics is that we are free from the danger of comparing paralogous genes, as may often be the case with nuclear genes (Jacobs et al. 1988b). We therefore consider that mitochondrial DNA data are applicable to the analysis of the relationship among ray-finned fishes, lungfishes, coelacanth, and tetrapods.

Materials and Methods

Preparation of Total DNA. Total DNA was isolated from the livers of a coelacanth *L.chalumnae* (kindly provided by the Japanese Mission for Scientific Research on the Coelacanth) and a lungfish *L. paradoxa* (obtained from South America) according to Sambrook et al. (1989) with a slight modification.

PCR Primers. According to the conserved amino acid sequences of COI, cytochrome oxidase subunit II, and NADH dehydrogenase subunit 2 genes for six species of vertebrates (human, Anderson et al. 1981; cow, Anderson et al. 1982; mouse, Bibb et al. 1981; rat, Gadaleta et al. 1989; chicken, Desjardins and Morais 1990; frog, Roe et al. 1985), two species of sea urchins (Jacobs et al. 1988a; Cantatore et al. 1989) and fruit fly (Clary and Wolstenholme 1985), PCR primers were synthesized with a DNA synthesizer (model 391, Applied Biosystems), as listed below:

L4987;5′-AAAAGGATCCGGNTGRGGNGGNYTNAAYGARAC-3′
L5947;5′-AAAAGGATCCTTY(AT)(GC)NACNAAYCAYAARGAYAT-3′
H6096;5′-AAAACTGCAGGYATNACYATRAARAARATYAT-3′
L6142;5′-AAAAGGATCCNATYATRATYGGNGGNTTYGGNAA-3′
H6270;5′-AAAACTGCAGGRTANACNGTYCANCCNGANCC-3′
L6565;5′-AAAAGGATCCCAAYYTNAAYTCNTCNTTYTTYGA-3′
H6606;5′-AAAACTGCAGTCNGGRTGNCCRAARAAYCARAA-3′
L6784;5′-AAAAGGATCCGTNTGRGCNCAYCAYATRTTYAC-3′
H7005;5′-AAAACTGCAGCNACNACRTARTANGARTCRTG-3′
L7192;5′-AAAAGGATCCAAYYTNACNTTYTTYCCNCARCAYTT-3′
H7212;5′-AAAACTGCAGGRTARTCN(GC)(AT)RTANCGNCGNGG-3′
H7886;5′-AAAACTGCAGTCRTAN(GC)(AT)YCARTAYCAYTGRTGNCC-3′

N indicates A, G, T, and C. R is A and G, and Y is T and C. The letters H and L indicate that the sequence of oligonucleotides is on the Heavy strand or Light strand, respectively. Numbers indicate the position of the 3′-end of the oligonucleotides, according to the human mitochondrial genome numbering system (Anderson et al. 1981). The underlined sequences indicate the introduced restriction sites to be inserted into cloning vectors. Some of them have already been reported in our previous paper (Yokobori et al. 1993).

Amplification and Sequencing of DNA Fragments. PCR was carried out according to Saiki (1989) with a slight modification. Conditions of amplification and sequencing of these fragments were the same as those described in our previous paper (Yokobori et al. 1993).

Data Set. Assuming that ray-finned fishes are an outgroup to the clade formed by sarcopterygians and tetrapods, we used the carp *Cyprinus carpio* (Chang and Huang 1991) and the loach *Crossostoma lacustre* (Tzeng et al. 1992) as outgroups in inferring a relationship among *L. chalumnae, L. paradoxa,* and the frog *Xenopus laevis* (Roe et al. 1985) as a representative of tetrapods. In the case of 12S rRNA, the sequences reported by Meyer and Dolven (1992) were used (as shown in Fig. 2 in their paper), and sites with any deletion or insertion were excluded from the analysis.

Phylogenetic Analyses. In analyses of the amino acid sequences of the COI and cyt b genes, we used a maximum likelihood (ML) method for inferring phylogenetic trees which was originally developed by Felsenstein (1981) for nucleotide sequences and later modified by Kishino et al. (1990) and Adachi and Hasegawa (1992) for amino acid sequences. Whereas the ML method, assuming no equal evolutionary rate among lineages, is robust against violation of the rate constancy among lineages, the maximum parsimony (MP) method widely used in molecular phylogenetics is not (Felsenstein 1978; Hasegawa et al. 1991). Although ML analysis does assume constancy of rates across amino acid positions, it has been shown that the method is robust to some extent against violation of this assumption (Hasegawa and Fujiwara 1993). Since ML analysis may depend upon the assumed model for amino acid substitutions during evolution, it is desirable to carry out the analysis using alternative models in order to examine the robustness of the results against violation of the assumed model.

Consequently, the following three Markov models for amino acid substitutions were used in the ML analysis of amino acid sequences of mitochondrial COI genes (Hasegawa et al. 1992a, 1993; Adachi et al. 1993): (1) the "Dayhoff model" that was developed by Kishino et al. (1990) and by Adachi and Hasegawa (1992), which assumes an empirical transition matrix compiled by Dayhoff et al. (1978); (2) the "Proportional model," in which it is assumed that, in an infinitesimally short time-interval *dt,* the probability $Pij(dt)$ of the current amino acid i being replaced by another amino acid j is given by

$$Pij(dt) = udt\pi j \tag{1}$$

where u is a parameter that determines the substitution rate and πj is the frequency of the amino acid j; and (3) the "Poisson model," in which it is assumed that $Pij(dt)$ is given by

$$Pij(dt) = udt \tag{2}$$

The Proportional model is an extension of Felsenstein's (1981) model for nucleotide sequences to amino acid sequences. The Poisson model, which is the simplest, has been used in most molecular phylogenetic studies, although it is sometimes not stated explicitly.

12S rRNA data were analyzed by the ML method for analyzing nucleotide sequences (DNAML in the PHYLIP package, version 3.2, Felsenstein 1989).

The maximum parsimony (MP) method and the neighbor-joining (NJ) method (Saito and Nei 1987) were also applied to the COI data. The MP analysis was carried out by the PROTRARS program in Felsenstein's (1989) PHYLIP package, version 3.2. To evaluate the extent to which the ML or MP tree is a significantly better representation of the true tree, we calculated standard errors (SE) of the differences of log-likelihoods and of substitution numbers of alternative trees, respectively, from those of the ML and the MP trees by the formula of Kishino and Hasegawa (1989), and estimated the bootstrap probability (the probability of being the ML or MP tree among alternative trees during bootstrap resampling; Felsenstein 1985) by the RELL (resampling of estimated log-likelihoods of sites or, for the MP method, resampling of estimated numbers of substitutions of sites) method given in the paper of Kishino et al. (1990) (10^4 replications). The RELL method is a good approximation to the computationally intensive bootstrap method (Felsenstein 1985) in estimating bootstrap probability (Hasegawa and Kishino 1994). For the NJ method, to estimate the bootstrap probability of each tree, resampling and tree reconstruction were replicated 500 times. The bootstrap probability of each tree was calculated from the occurrence frequency.

Results and Discussion

Amplification of the COI Gene by PCR

DNA fragments derived from COI genes of the coelacanth and the lungfish were amplified by using the polymerase chain reaction (PCR) (Saiki et al. 1988). Six fragments which partially overlapped each other and covered the whole COI gene region were obtained. The length of each fragment was identical to those estimated from the COI gene sequence of the frog, *X. laevis* (Roe et al. 1985). The nucleotide sequences of these fragments were determined and found to cover the mitochondrial genome ranging from the $tRNA^{Cys}$ gene to the $tRNA^{Asp}$ gene (almost 2 kbp). The positions of the initiation and termination codons were estimated from the locations of the $tRNA^{Tyr}$ and $tRNA^{Ser}_{UGA}$ genes, respectively, both of which are adjacent to the COI gene in the vertebrate mitochondrial genomes. The length of the *L. chalumnae* COI gene was estimated to be 1,545 bp (515 amino acids), and that of the *L. paradoxa* COI gene 1,551 bp (517 amino acids).

Amino Acid Sequence of COI Genes

The amino acid sequences of the COI genes were deduced from their nucleotide sequences according to the vertebrate mitochondrial genetic code (Osawa et al. 1992). Since the amino acid sequences of COI are highly conserved among vertebrates, the alignment of amino acid sequences in almost all regions does not cause ambiguity, and 504 amino acids of COI were provided for phylogenetic analyses (Fig. 2) by means of the ML, MP, and NJ methods. The estimated initiation codons of COI genes of the coelacanth and the lungfish are GTG (for valine), like those for the ray-finned fishes, the carp, *C. capio* (Chang and Huang 1991) and the loach, *C. lacustre* (Tzeng et al. 1992), which are used as outgroups in this analysis, but different from that of the frog COI gene (ATG, for methionine) (Roe et al. 1985). Thus, the N-terminal amino acid of COI was excluded here. The C-terminal region downstream from the 506th residue was also omitted, because alignment of this region is ambiguous.

```
          2
Coelacanth MITRWLFSTNHKDIGTLYMIFGAWAGMVGTALSLLIRAELSQPGALLGDDQIYNVVVTAH
  Lungfish T*****************L*******************************F**L****
     Frog A****************LV******L***************T*********I****
      Carp A****F***********LV**********************S**S*******I****
     Loach A****F***********LV********************N*************I****

         62
          AFVMIFFMVMPIMIGGFGNWLIPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLACSGVYAG
          ************************************L*********AF*****G***E**
          **I*****************V************************F*****S***E**
          ************L********V************************F*****S***E**
          ************L********V*******H*****************F*****S***E**

        122
          AGTGWTVYPPLAGNLAHAGASVDLTIFSLHLAGVSSILGAINFITTVINIKPPTMTQYQT
          *********************************M****DP*********M***AAS*F**
          ************************************I***********T**M***A*S****
          *****S*****************************************T**M***AIS****
          ********************************************T**M***ALS****

        182
          PLFIWSVLVTAVLLLLSLPVLAAGITMLLTDRNLNTTFFVPVGGGDPILYQHLFWFFGHL
          ****C**MI*T*****************************D*A*****************P
          ***V****I*******************************D*A*****V***********P
          ***V*************************************D*A*****************P
          ***V*A**I*******************************D*A*****************P

        242
          EVYILILPGFGMISHIVAYYSGKKEPFGYMGMVWAMMATGLLGFIVWAHHMFTVGMDVDT
          *****************T******************I*I********************
          *****************T********************SI***************DLN***
          ***********I***V*********************I*********************
          ***********I***V****A****************I*********************

        302
          RVYFTSATMIIAIPTGVKVFSWLATLHGGVTKWDTPLLWALGFIFLFTVGGLTGIVLANS
          *A**************************AI**E***I****************V*****
          *A***********************M***TI***A*M**********************
          *A**************************SI**E**M********************S**
          *A*************************TI*****M********************S**

        362
          SLDIILHDTYYVVAHFHYVLSMGAVFAIMGGLVHWFPLMTGYTLHNTWTKIHFGVMFTGV
          ****M************************A***************D****M****T*I**
          ****M*****************************FI*****F******E**A********A**
          ****V************************AAF******L******S***********I**
          ****V************************A*F******F**FS**D***********I**

        422
          NLTFFPQHFLGLAGMPRRYSDYPDAYTLWNTVSSIGSLISLIAVIMFMFILWEAFLAKRE
          ***************************F******V**V**MV**MLLL**I****AS***
          ************SA*****************************V****M***I****A****
          ************SA************A*************V*****L*******A****
          ****************************************V***I*L*******AS**Q

        482
          VLIVEMTTTNVEWLHGCPPPHHTY
          *SSI***H**A*********Y**F
          *TTY*L*S*ML***Q***T*Y**L
          **S**L*A***********Y***
          *MS**L*M***********Y**F
```

Fig. 2. Alignment of amino acid sequences of COI of various animal mitochondria. Amino acids are indicated with *one-letter codes*. The first position and the C-terminal region downstream from the 506th residue, which are not used in this analysis, are now shown. 1st column, coelacanth (*L. chalumnae*); 2nd, lungfish (*L. paradoxa*); 3rd, *Xenopus* (*X. laevis*) (Roe et al. 1985); 4th, carp (*C. carpio*) (Chang and Huang 1991); 5th, loach (*C. lacustre*) (Tzeng et al. 1992).

Table 1. Analyses of COI amino acid sequences by three models of the ML method for inferring protein phylogeny, and by the MP method[a]

Tree topology	ML			MP
	Dayhoff	Prop.	Poisson	
1. (Coe, (Lun, Tet))	−11.5 ± 8.9	−19.0 ± 11.0	−21.6 ± 12.4	+7 ± 4.1
2. (Lun, (Coe, Tet))	−13.2 ± 8.4	−19.0 ± 11.0	−21.7 ± 12.4	+8 ± 4.0
3. ((Coe, Lun), Tet)	ML	ML	ML	MP

[a] Numbers in the table show the difference of log-likelihood of each tree from that of the ML tree and the difference of the substitution number of each tree from that of the MP tree. ± indicates standard error. Coe: coelacanth, *L. chalumnae;* Lun: lungfish, *L. paradoxa;* Tet: frog, *X. laevis*

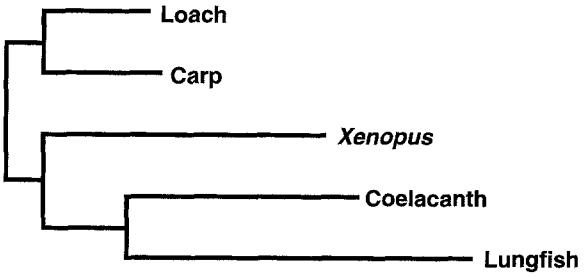

Fig. 3. Maximum likelihood tree of the relationship among the coelacanth, lungfish, frog, carp, and loach, inferred by the ML method with the Dayhoff model using COI amino acid sequences. Coelacanth (*L. chalumnae*), lungfish (*L. paradoxa*), *Xenopus* (*X. laevis*) (Roe et al. 1985), carp (*C. carpio*) (Chang and Huang 1991), loach (*C. lacustre*) (Tzeng et al. 1992). The lengths of horizontal branches indicate the relative evolutionary rate of each branch.

Phylogenetic Analyses of Amino Acid Sequences of COI Genes

We analyzed the COI genes by the ML method to infer the protein phylogeny (PROTML) (Adachi and Hasegawa 1992). Only amino acid sequences were used in the analyses, because a more realistic model is available for the amino acid substitution process than for the nucleotide substitution process in protein-encoding genes (Adachi and Hasegawa 1992; Adachi et al. 1993).

The result of the ML analysis by the Dayhoff model is shown in Table 1. This indicates that Tree-3, exhibiting the coelacanth/lungfish clade, is the maximum likelihood tree, as shown in Fig. 3. The relative substitution rates of the coelacanth and the lungfish are high in the COI amino acid sequences as compared with those of the ray-finned fishes, the carp, and the loach.

The ML analyses by the Proportional and Poisson models also consistently support Tree-3 (Table 1), showing a coelacanth/lungfish clade irrespective of the model for the amino acid substitution (Table 1).

Bootstrap analysis is most commonly used to estimate the reliability of a phylogenetic tree (Felsenstein 1985). We estimated the approximate bootstrap probabilities of each tree in Fig. 1 for the ML analyses based on the three models by using the RELL method. The bootstrap probability of Tree-3 is higher than 90% in all models, while Tree-2, with a coelacanth/tetrapod clade, is unlikely in any of the models (Table 2). However, since log-likelihood differences of Tree-1 and Tree-2 from Tree-3 do not exceed 2SE (Table 1), these trees cannot be completely ruled out from analysis.

Furthermore, both the MP and the NJ analyses are congruent with those by ML. The most parsimonious tree is Tree-3, and the NJ method also gives Tree-3. The bootstrap probability of Tree-3 is higher than 90% in the MP and NJ methods, as suggested by the ML analyses. Thus, although Tree-1, with a lungfish/tetrapod clade, cannot be completely excluded because the bootstrap probability of Tree-1 by ML analysis with the Dayhoff model is higher than 5%, Tree-3 is preferentially supported by the COI data.

Phylogenetic Analysis of Mitochondrial Genes

How are our new data compatible with previous data supporting a lungfish/tetrapod clade?

To obtain a consensus among the mitochondrial genes, we reanalyzed the 12S rRNA data (Meyer and Wilson 1990; Meyer and Dolven 1992) and the cyt b data (Meyer and Wilson 1990) by the ML method and compared them with the results for the COI data (Table 3). In the case of 12S rRNA, the data set in Meyer and Dolven (1992) was used. In the case of cyt b, the carp and loach were used as the outgroup, referring to the data of lungfish and coelacanth reported by Meyer and Wilson (1990). Consistent with the findings of the previous authors, Tree-1 is supported by the 12S rRNA and cyt b data. However, the COI data are not necessarily contradictory to those of 12S rRNA and cyt b, because the latter data do not necessarily rule out Tree-3. The ML method has an advantage in that it enables the results of tree topologies inferred from diverse genes to be synthesized (the bottom row in Table 3). After separate branch lengths were estimated for each gene, the sampling of sites by the RELL method was done separately within each gene. The overall evidence for the mitochondrial genes rules out Tree-2 (a bootstrap probability of 0.002) and supports Tree-1 almost equally with Tree-3 (Table 3).

Although the bootstrap probabilities of Tree-1 and Tree-3 with respect to the overall evidence for the mitochondrial genes are equal (0.50, respectively), the ev-

Table 2. Bootstrap probabilities of ML, MP, and NJ trees estimated from amino acid sequences of the COI gene[a]

	ML				
Tree topology	Dayhoff	Prop.	Poisson	MP	NJ
1. (Coe, (Lun, Tet))	0.0741	0.0253	0.0208	0.0281	0.045
2. (Lun, (Coe, Tet))	0.0243	0.0271	0.0257	0.0073	0.033
3. ((Coe, Lun), Tet)	0.9016	0.9476	0.9535	0.9646	0.922

[a] Bootstrap probabilities of each tree were estimated by the RELL method (Kishino et al. 1990) with a sample size of 10^4. For the NJ method of estimation of the bootstrap probability of each tree, resampling and tree reconstruction were replicated 500 times. The probability of each tree is calculated from the occurrence frequency. Coe: coelacanth, *L. chalumnae;* Lun: lungfish, *L. paradoxa;* Tet: frog, *X. laevis*

Table 3. ML analyses of 12S rRNA, cyt b, and COI encoded by mitochondrial DNA[a]

Gene (length)	Tree-1 (Coe, (Lun, Tet))	Tree-2 (Lun, (Coe, Tet))	Tree-3 ((Coe, Lun), Tet)
12S rRNA (234 bp)	ML	−6.46 ± 4.43	−5.86 ± 4.79
cyt b (120 aa)	ML	−5.87 ± 4.94	−5.68 ± 5.07
COI (504 aa)	−11.47 ± 8.90	−13.20 ± 8.44	ML
Total	ML	−14.05 ± 7.88	−0.07 ± 11.29
	(0.50)	(0.00)	(0.50)

[a] 12S rRNA was analyzed by DNAML in PHYLIP version 3.2 (Felsenstein 1989) and cyt b and COI were analyzed by the ML method based on the Dayhoff model (Kishino et al. 1990; Adachi and Hasegawa 1992). The consensus of bootstrap probabilities of each tree was estimated by the RELL method (Kishino et al. 1990) with a sample size of 10^4; these are shown in parentheses

idence of the 12S rRNA data for Tree-1 may be discounted for the following reasons. First, the base composition of rRNA is sensitive to directional mutation pressure (Osawa et al. 1992), whereas the amino acid composition of a conservative protein is not significantly affected by the pressure, as compared with the nucleotide sequences (Hasegawa et al. 1992b; Hasegawa and Hashimoto 1993). Second, since the base substitution in a base-paired stem region of rRNA occurs with correlation, an actual error would be larger than that in Table 3 estimated by assuming independence among sites. Third, the alignment of rRNA is not free from ambiguity. Of course, the cyt b data preferentially support Tree-1, as do the 12S rRNA data. And in the case of cyt b, the data would be free from the problems referred to above with respect to the rRNA data. Here, we can only point out that the possibility of the coelacanth/lungfish clade is large enough not to be excluded from the search for the origin of tetrapods, although it has been the minority hypothesis.

Recently, Hedges et al. (1993) reported the phylogenetic analysis of mitochondrial genes, 12S rRNA, $tRNA^{Val}$, 16S rRNA, and cyt b (almost 3 kbp, in all) of carp, coelacanth, lungfishes, and tetrapods. However, because their data are not presented in a form suitable for our analysis, we are unable to include them in this study. Hedges et al. (1993) claim that their analysis data support the lungfish/tetrapod clade (Tree-1) as do the data of Meyer and Wilson (1990) and the Meyer and Dolven (1992).

Phylogenetic Analyses of Nonmitochondrial Molecular Data

Gorr et al. (1991) concluded from a phylogenetic analysis of the amino acid sequence of hemoglobins that coelacanths and tetrapods are monophyletic and lungfishes are their sister group (Tree-2). However, the results of reanalyses of their data (Forey 1991; Stock and Swofford 1991; Sharp et al. 1991; Meyer and Wilson 1991) indicated that hemoglobin data cannot clarify the relationship among coelacanths, lungfishes, and tetrapods. Hasegawa reanalyzed the hemoglobin data using the ML method (unpublished data) and showed that the β hemoglobin data only weakly supported the coelacanth/tetrapod clade (Tree-2), as Gorr et al. (1991) did. Moreover, the α hemoglobin data also support the coelacanth/lungfish clade only slightly (Tree-3). The serious problem with hemoglobin data is that the distinction between orthologous and paralogous relations is not clear (Forey 1991); 18S rRNA data (Stock et al. 1991) could not clarify this relationship.

Other molecular data, such as those of parvalbumin (Maeda et al. 1984) and 23S rRNA (Hillis et al. 1991), are not available for the lungfish. Nomark et al. (1991) analyzed the relationship of fishes from the partial sequences of mitochondrial genes, COI, COII, and cyt b. Although the COI and COII data were not available for the lungfish and the coelacanth, cyt b data were analyzed for various fishes, including an African lungfish, *Protopterus* sp., in addition to the work of Meyer and

Wilson (1990). The lungfish/tetrapod clade was supported by these cyt b analyses.

Morphological Data Supporting the Coelacanth/lungfish Clade

A coelacanth/lungfish clade has not been widely accepted in previous studies, though some morphologists have reached this conclusion. Some lines of evidence supporting Tree-3—namely, the coelacanth/lungfish clade—are shown below.

By analyzing morphological elements of neural systems, Northcutt (1987) described common primitive characteristics between the coelacanth *L. chalumnae* and the Australian lungfish *Neoceratodus forsteri,* the most primitive extant lungfish (neural characters 12–16 in Fig. 10, p. 293 of his paper) (Northcutt 1987). These characteristics are not shared by the amphibians and the Lepidosirenoids (including *L. paradoxa*), and there are similarities between the amphibians and the Lepidosirenoids. But the characteristics shared by amphibians and Lepidosirenoids would have independently occurred in each of the lineages, because Lepidosirenoids are more adapted to the amphibious lifestyle than *N. forsteri.* The extant lungfishes are thought to be a monophyletic group (Forey 1987). Therefore, Northcutt believes that the similarities of the primitive forms between the coelacanth and the Australian lungfish indicate their close relationship.

Chang (1991) describes several characteristics (indicated by characters 7 and 9–11 in her paper) of bone morphology shared by coelacanths and lungfishes, but not by tetrapods. Character 9 in the paper was cited as a sarcopterygian synapomorphy subsequently lost in tetrapods in the papers of Rosen et al. (1981) and Schultze (1987).

In addition, Marshall and Schultze (1992) pointed out the following: although Meyer and Wilson (1990) drew attention to several common characteristics shared by the lungfishes and the coelacanths as a result of parallelisms or reversals under the scheme of the lungfish/tetrapod clade [traits 15–21 of Table 2 in Meyer and Wilson (1990)], these could be considered to support Tree-3—namely, the coelacanth/lungfish clade. Marshall and Schultze (1992) also pointed out a possibility that, in the traits cited by Meyer and Wilson (1990) supporting Tree-1 (traits 1–14 in Table 2 in their paper), only one characteristics (trait 14) would support the lungfish/tetrapod clade (Tree-1).

Conclusion

Because radiation among lungfishes, coelacanths, and tetrapods is thought to have occurred during a short period in the Devonian period (Forey 1988; Marshall and Schultze 1992), it might be difficult to resolve their branching order. Therefore it seems important to evaluate their relationship not only by paleontological, physical, and morphological data, but also by molecular data.

The coelacanth/lungfish clade (Tree-3) has so far received less attention than the alternatives (Tree-1 and -2). In this study, we have demonstrated that the minority view suspecting Tree-3 to be correct should not be dismissed, and that the traditional view supporting the coelacanth/tetrapod clade (Tree-2) should be ruled out. However, we could not determine whether the coelacanth/lungfish or lungfish/tetrapod clade reflects the true relationship among sarcopterygians and tetrapods. The scenario of tetrapod evolution needs to be reexamined in the future using more sequence data.

Acknowledgments. We are indebted to Dr. T. Yano of the College of Arts and Sciences, Showa University, for valuable discussions. We also thank the Japanese Mission for Scientific Research on the Coelacanth for providing us with the liver of a coelacanth, *L. chalumnae.* This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas for the Ministry of Education, Science and Culture of Japan.

References

Adachi J, Hasegawa M (1992) Computer science monographs no. 27, MOLPHY: Programs for molecular phylogenetics I — PROTML: Maximum likelihood inference of protein phylogeny. Institute of Statistical Mathematics, Tokyo

Adachi J, Cao Y, Hasegawa M (1993) Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: rapid evolution in warm-blooded vertebrates. J Mol Evol 36:270–281

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG (1982) The complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717

Bibb MJ, van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol 18:225–239

Cantatore P, Roberti M, Rainaldi G, Gadaleta MN, Saccone C (1989) The complete nucleotide sequence, gene organization and genetic code of mitochondrial genome of *Paracentroutus lividus.* J Biol Chem 264:10695–10975

Chang M-M (1991) "Rhipidistians", dipnoans, and tetrapods. In: Schultze H-P, Trueb L (eds) Origins of the higher groups of tetrapods. Controversy and consensus. Cornell University Press, Ithaca, NY, pp 3–28

Chang YS, Huang FL (1991) EMBL accession number X61010

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba:* nucleotide sequence, gene organization, and genetic code. J Mol Evol 22:252–271

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington, DC, pp 345–352

Desjardins P, Morais R (1990) Sequence and organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. J Mol Biol 212:599–634

Felsenstein J (1978) Cases in which parsimony and compatibility methods will be positively misleading. Syst Zool 27:401–410
Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376
Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
Felsenstein J (1989) PHYLIP, Version 3.2. University of Washington, Seattle
Forey PL (1987) Relationships of lungfishes. J Morphol [suppl]1:75–91
Forey PL (1988) Golden jubilee for the coelacanth *Latimeria chalumnae*. Nature 336:727–732
Forey PL (1991) Blood lines of the coelacanth. Nature 351:347–348
Fritsch B (1987) Inner ear of the coelacanth fish *Latimeria* has tetrapod affinities. Nature 327:153–154
Gadaleta G, Pepe G, Decandia G, Quagliariello C, Sbisa E, Saccone C (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 28:497–516
Gorr T, Kleinschmidt T, Fricke H (1991) Close tetrapod relationships of the coelacanth *Latimeria* indicated by haemoglobin sequences. Nature 351:394–397
Hasegawa M, Cao Y, Adachi J, Yano T (1992a) Rodent polyphyly? Nature 355:595
Hasegawa M, Fujiwara M (1993) Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny. Mol Phylogenet Evol 2:1–5
Hasegawa M, Hashimoto T (1993) Ribosomal RNA trees misleading? Nature 361:23
Hasegawa M, Hashimoto T, Adachi J (1992b) Origin and evolution of eukaryotes as inferred from protein sequence data. In: Hartman H, Matsumoto K (eds) The origin and evolution of the cell. World Scientific, Singapore, pp 107–130
Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T (1993) Early branchings in the evolution of eukaryotes: ancient divergence of Entamoeba that lacks mitochondria revealed by protein sequence data. J Mol Biol 36:380–388
Hasegawa M, Kishino H (1994) Accuracies of the simple methods for estimating the bootstrap probability of a maximum-likelihood tree. Mol Biol Evol (in press)
Hasegawa M, Kishino H, Saitou N (1991) On the maximum likelihood method in molecular phylogenetics. J Mol Evol 32:443–445
Hedges SB, Hass CA, Maxon LR (1993) Relations of fish and tetrapods. Nature 363:501–502
Hennig W (1983) Stammesgeschichte der Chordaten. P Parey, Hamburg
Hillis DM, Dixon MT, Ammerman LK (1991) The relationships of the coelacanth *Latimeria chalumnae:* evidence from sequences of vertebrate 28S ribosomal RNA genes. Environ Biol Fishes 32:119–130
Irwin DM, Kocher KD, Wilson AC (1990) Evolution of the cytochrome b gene of mammals. J Mol Evol 32:128–144
Jacobs HT, Elliot DJ, Math VB, Farquharson A (1988a) Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J Mol Biol 202:185–217
Jacobs HT, Balfe P, Cohen BL, Farquharson A, Comito L (1988b) Phylogenetic implications of genome rearrangement and sequence evolution in echinoderm mitochondrial DNA. In: Paul CRC, Smith AB (eds) Echinoderm phylogeny and evolutionary biology. Clarendon Press, Oxford, pp 121–137
Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol 29:170–179
Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood of protein phylogeny and the origin of chloroplasts. J Mol Evol 31: 151–160
Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A 86:6196–6200
Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J Mol Evol 37:380–398
Maeda N, Zhu D, Fitch WM (1984) Amino acid sequence of lower vertebrate parvalbumins and their evolution. Mol Biol Evol 1:473–488
Marshall C, Schultze H-P (1992) Relative importance of molecular, neontological, and paleontological data in understanding the biology of the vertebrate invasion of land. J Mol Evol 35:93–101
Meyer A, Wilson AC (1990) Origins of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J Mol Evol 31:359–364
Meyer A, Wilson AC (1991) Coelacanth's relationships. Nature 353: 219
Meyer A, Dolven SI (1992) Molecules, fossils, and the origin of tetrapods. J Mol Evol 35:102–113
Miles R (1975) The relationships of the Dipnoi. Collq Int CNRS 218:133–148
Miles R (1977) Dipnoan (lungfish) skulls and the relationships of the group: a study based on new species from the Devonian of Australia. Zool J Linn Soc 61:1–328
Miyamoto MM, Boyle SM (1989) The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 437–450
Normark BB, McCane AR, Harrison RG (1991) Phylogenetic relationships of neopterygian fishes, inferred from mitochondrial DNA sequences. Mol Biol Evol 8:819–834
Northcutt RG (1987) Lungfish neural characters and their bearing on sarcopterygian phylogeny. J Morphol [Suppl]1:277–297
Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. J Microbiol Rev 56:229–264
Roe BA, MA DP, Wilson RK, Wong JF-H (1985) The complete sequence of the *Xenopus laevis* mitochondrial genome. J Biol Chem 260:9759–9774
Romer AS (1966) Vertebrate paleontology, 3rd ed. University of Chicago Press, Chicago
Rosen DE, Forey PL, Gardiner BG, Patterson C (1981) Lungfishes, tetrapods, paleontology and plesiomorphy, Bull Am Mus Nat Hist 167:159–276
Saiki RK, Gelfand DH, Stoffel S, Schorf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491
Saiki RK (1989) The design and optimization of the PCR. In: Erlich HA (ed) PCR technology, Stockton Press, NY, pp 7–16
Saito N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Schultze H-P (1987) Dipnoans as Sarcopterygians. J Morphol [Suppl]3:39–74
Schultze H-P, Campbell KSW (1987) Characterization of the Dipnoi, a monophyletic group. J Morphol [Suppl]1:25–37
Sharp PM, Lloyd AT, Higgins DG (1991) Coelacanth's relationship. Nature 353:218–219
Stock DW, Moberg KD, Maxson LR, Whitt GS (1991) A phylogenetic analysis of the 18S ribosomal RNA sequence of the coelacanth *Latimeria chalumnae*. Environ Biol Fishes 32:99–117
Stock DW, Swofford DL (1991) Coelacanth's relationships. Nature 353:217–218
Tzeng C-S, Hui C-F, Shen S-C, Huang PC (1992) The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. Nucleic Acids Res 20:4853–4858
Yokobori S, Ueda T, Watanabe K (1993) Codons AGA and AGG are read as glycine in ascidian mitochondria. J Mol Evol 36:1–8