Complete Mitochondrial Genome Sequences of the South American and the Australian Lungfish: Testing of the Phylogenetic Performance of Mitochondrial Data Sets for Phylogenetic Problems in Tetrapod Relationships

Henner Brinkmann,[1]* Angelika Denk,[1]† Jürgen Zitzler,[1]‡ Jean J. Joss,[2] Axel Meyer[1]

[1] Department of Biology, University of Konstanz, 78457 Konstanz, Germany
[2] Department of Biological Sciences, Macquarie University, New South Wales 2109 Australia

Abstract. We determined the complete nucleotide sequences (16403 and 16572 base pairs, respectively) of the mitochondrial genomes of the South American lungfish, *Lepidosiren paradoxa*, and the Australian lungfish, *Neoceratodus forsteri* (Sarcopterygii, Dipnoi). The mitochondrial DNA sequences were established in an effort to resolve the debated evolutionary positions of the lungfish and the coelacanth relative to land vertebrates. Previous molecular phylogenetic studies based on complete mtDNA sequences, including only the African lungfish, *Protopterus dolloi*, sequence were able to strongly reject the traditional textbook hypothesis that coelacanths are the closest relatives of land vertebrates. However, these studies were unable to statistically significantly distinguish between the two remaining scenarios: lungfish as the closest relatives to land vertebrates and lungfish and coelacanths jointly as their sister group (Cao et al. 1998; Zardoya et al. 1998; Zardoya and Meyer 1997a). Lungfish, coelacanths, and the fish ances-

tors of the tetrapod lineage all originated within a short time window of about 20 million years, back in the early Devonian (about 380 to 400 million years ago). This short divergence time makes the determination of the phylogenetic relationships among these three lineages difficult. In this study, we attempted to break the long evolutionary branch of lungfish, in an effort to better resolve the phylogenetic relationships among the three extant sarcopterygian lineages. The gene order of the mitochondrial genomes of the South American and Australian lungfish conforms to the consensus gene order among gnathostome vertebrates. The phylogenetic analyses of the complete set of mitochondrial proteins (without ND6) suggest that the lungfish are the closest relatives of the tetrapods, although the support in favor of this scenario is not statistically significant. The two other smaller data sets (tRNA and rRNA genes) give inconsistent results depending on the different reconstruction methods applied and cannot significantly rule out any of the three alternative hypotheses. Nuclear protein-coding genes, which might be better phylogenetic markers for this question, support the lungfish–tetrapod sister-group relationship (Brinkmann et al. 2004).

Present address: Département de biochimie, Université de Montreal, Montreal, QC H3C3J7, Canada
†*Present address*: Max Planck Research Center for Ornithology, Postfach 1564, 82305 Starnberg, Germany
‡*Present address*: Xantos biomedicine AG, Max-Lebsche Platz 31, 81337 Munich, Germany

Correspondence to: Axel Meyer; *email*: axel.meyer@uni-konstanz.de

Table 1. Systematic position of the lungfish[a]

Class: Osteichthyes (bony fish)
  Subclass: Actinopterygii (ray-finned fish)
    Chondrostei (sturgeon, *Acipenser*; birchir, *Polypterus*)
    Neopterygii (gar, *Lepisosteus*; bowfish, *Amia*)
    Teleostei (modern ray-finned fish)
  Subclass: Sarcopterygii (lobe-finned fish)
    Actinistia (coelacanth, *Latimeria*)
    Rhipidistia
      Dipnoi (lungfish)
      Ceratodontiformes
        Ceratodontidae (*Neoceratodus*, Australian lungfish)
      Lepidosireniformes
        Lepidosirenidae (*Lepidosiren*, South American lungfish)
        Protopteridae (*Protopterus*, African lungfish)
      Porolepiformes[b]
      Osteolepiformes[b]
      Tetrapoda (land vertebrates)
        Lissamphibia (modern amphibians)
        Amniota (reptiles, birds, mammals)

[a]Modified from Zardoya and Meyer (1996a).
[b]Extinct.

Introduction

The conquest of terrestrial habitats by previously water-living vertebrates in the Devonian approximately 360 million years ago (mya) was one of the most important evolutionary events in the history of vertebrates (reviewed in Benton 1990; Zimmer 1999, p 290). Yet the important question about the identity of the closest living relative among fish and land vertebrates is still not fully resolved (reviewed in Meyer 1995; Zardoya et al. 2003; Zardoya and Meyer 2001c), at least based on an all molecular data set (Meyer and Zardoya 2003). Many morphological and physiological changes had to occur to permit the conquest of land by vertebrates of aquatic origin (e.g., the modification of swimming limbs into load-bearing limbs and alteration of the respiratory as well as the osmoregulatory systems (reviewed in, e.g., Panchen 1987) to make life on land possible. The identification of the closest living relative among fish to land vertebrates might permit a better understanding and reconstruction of the necessary preadaptations that permitted the colonization of land (Meyer and Dolven 1992).

It is widely accepted that lobe-finned fish (Sarcopterygii), which include the coelacanths (Actinista), lungfish (Dipnoi), and several extinct rhipidistian groups as well as the extant tetrapods, form a monophyletic group (Table 1). Among the sarcopterygians, one of the extinct lineages of rhipidistians fish (Elpistostegids, represented by the fossil *Panderichtys*), but not representatives of an extant clade, seems to be the closest relative of land vertebrates (Ahlberg et al. 1996; Cloutier and Ahlberg 1996). Although lobe-finned fish were a highly successful

and species-rich group during the lower Devonian (400 mya), only six lungfish (Carroll 1988; Marshall 1987) and one or two coelacanth species (Holder et al. 1999; Pouyaud et al. 1999) survived until today. These "living fossils" are of special interest to several biological disciplines (e.g., developmental biology, comparative morphology, and paleontology) since their morphology, physiology, neurobiology, and biochemistry may still correspond, or at least be similar, to some of the conditions present in the last common ancestor of all land vertebrates (Meyer and Dolven 1992; Meyer and Wilson 1990; Zardoya and Meyer 1996b). The relationships among the three living Sarcopterygiian groups (coelacanth, lungfish, and tetrapods) remain, however, somewhat ambiguous based on phenotypic (morphological, paleontological, neurobiological) (Ahlberg and Milner 1994; Forey 1998a; Patterson 1980; Rosen 1981) and genetic data (for reviews see Meyer 1995; Meyer and Zardoya 2003; Zardoya and Meyer 2001, 2003; Kikugawa et al. 2004; Takezaki et al. 2004).

Lungfish were discovered over 150 years ago (Bischoff 1840), and based on several adaptations such as lungs and estivation (which allows lungfish to survive droughts) and reduced scales and fins, these obligate air breathing fish were initially believed to belong to the amphibians. Extant lungfish are strictly limited to freshwater and are found in Australia (*Neoceratodus forsteri*), Africa (four species of the genus *Protopterus*), and South America (*Lepidosiren paradoxa*). This distribution of the living lungfish suggests that the age of the two orders Ceratodontiformes (*Neoceratodus*) and Lepidosireniformes, with two families (*Protopterus* and *Lepidosiren*), predates the split of the Gondwana supercontinent into the current continents more than 180 mya (for systematic position see also Table 1).

The other extant group of lobe-finned fish (besides the tetrapods), the coelacanths, were long believed to have gone extinct about 80 mya until the spectacular discovery of a surviving specimen off the coast of East Africa (Smith 1939a, b). Similarly exciting was the recent finding of another new species of coelacanth off the coast of Sulawesi in Indonesia (Erdmann 1999; Erdmann et al. 1999; Forey 1998b; Pouyaud et al. 1999). Until today, the coelacanth is depicted in many biological textbooks as the "missing link" between sarcopterygian fish and land vertebrates (e.g., Campbell 1987; Romer 1966). However, most molecular and recent morphological and paleontological studies suggest that the long-time favored coelacanth + tetrapod sister-group relationship is the most unlikely of the three possible scenarios (Fig. 1) and can be statistically rejected based on large molecular data sets (Cao et al. 1998; reviewed in Meyer and Zardoya 2003; Zardoya et al. 1998; Zardoya and Meyer 2001c). Until now, neither
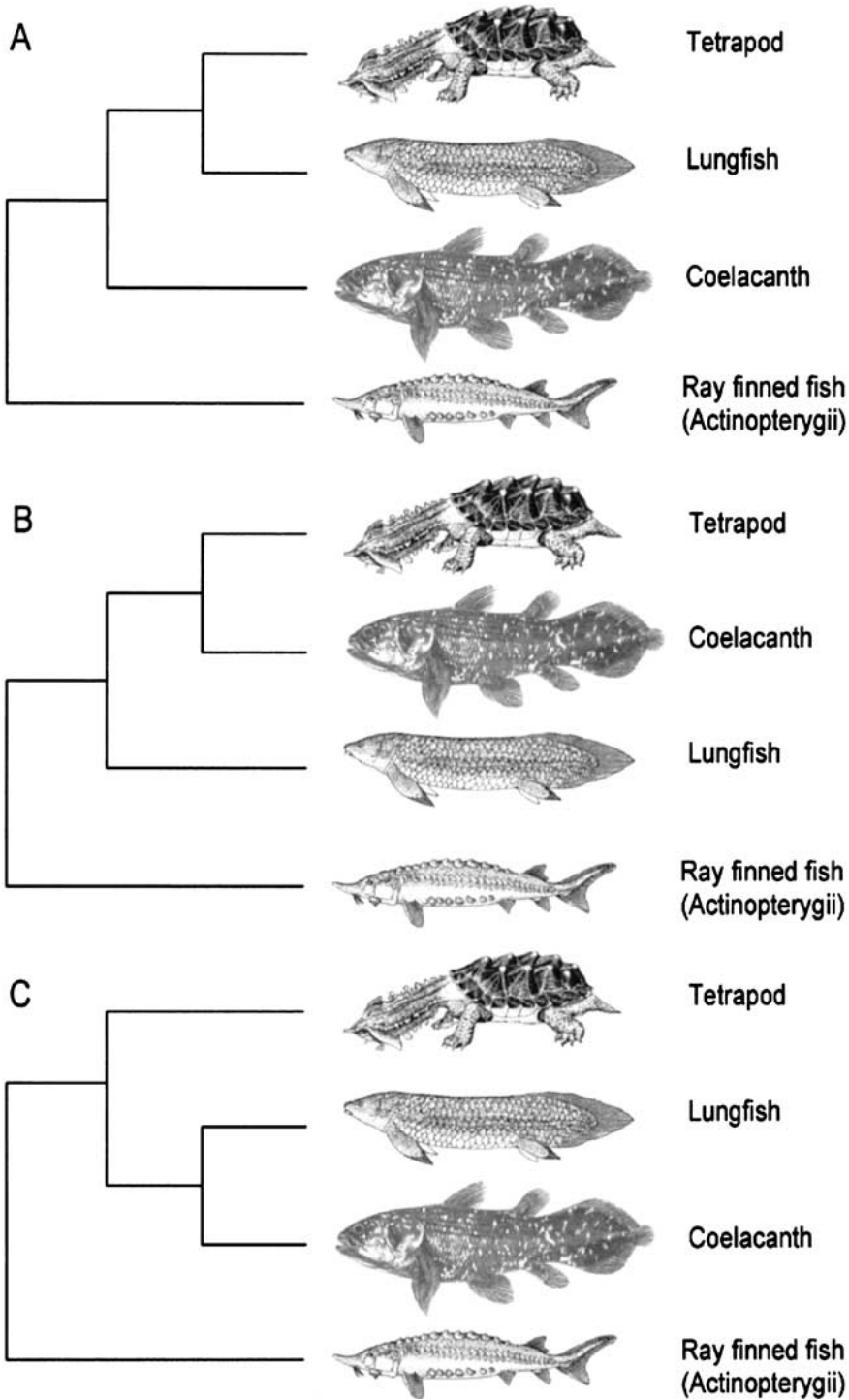
Fig. 1. Alternative hypotheses of sister-group relationships between Sarcopterygii and tetrapods.
A Lungfish as closest living relative of tetrapods. **B** Coelacanth as sister group to tetrapods. **C** Coelacanth and lungfish form a monophyletic sister group of tetrapods.

of the two remaining hypotheses among the sarcopterygian lineages (lungfish + tetrapod versus lungfish + coelacanth sister-group relationships) could be ruled out significantly based on mitochondrial DNA data (Fig. 1). In 1999, the complete mitochondrial DNA genomes of at least 33 land living vertebrates (amphibians, reptiles including birds, and placental and marsupial mammals) have been reported (Curole and Kocher 1999); today there are already 10 times

more (~331), yet only one lungfish and one coelacanth complete mtDNA sequence is known.

Several researchers demonstrated a relatively poor performance of individual mitochondrial genes in recovering phylogenetic relationships among lineages that originated in the Devonian (Cummings et al. 1995; Russo et al. 1996; Zardoya and Meyer 1996c). The phylogenetic reliability of combined mitochondrial data sets from protein, tRNA, and rRNA genes

is clearly higher than that of any single or partial mitochondrial data set (Cummings et al. 1995; Russo et al. 1996; Zardoya and Meyer 1996c). This is especially true for the largest data set, consisting of the combined mitochondrial protein sequences (Zardoya and Meyer 1996a,c)

Unbalanced species sampling and variable rates of molecular evolution in different lineages can make phylogenetic reconstruction prone to errors. In contrast to the diverse and relatively well-represented tetrapods, the two remaining sarcopterygian lineages (lungfish and coelacanths) were only represented by a single sequence each in previous molecular phylogenetic studies. This created long branches for these species-poor ancient lineages, which diverged almost 400 mya. Old lineages might be saturated in terms of DNA sequence divergence, and if, in addition, they also have pronounced differences in evolutionary rates, phylogenetic reconstruction becomes challanging. Long, uninterrupted branches are known to potentially lead to artificial groupings in phylogenetic analyses (Brinkmann and Philippe 1999; Felsenstein 1978; Philippe et al. 2000), and "long-branch attraction" is therefore also a potentially important phylogenetic issue for the origin of tetrapods problem. Since until recently only a single coelacanth species was known to exist, and the newly discovered coelacanth species is very closely related (Holder et al. 1999; Brinkmann et al. unpublished data), nothing can be done about this. However, there are three extant lineages of lungfish of rather ancient origin, of which only the African lungfish lineage has been sampled before (Zardoya and Meyer 1996a).

Here we attempt to reduce the potentially negative effect of a long unbroken lungfish branch on the phylogenetic reconstruction by (1) a more complete species sampling through the addition of two lungfish sequences and (2) through phylogenetic emphasis on more slowly evolving and, thereby, probably phylogenetically more reliable sites. In the case of the protein coding genes (amino acid data) a reduction of phylogenetic noise in the data was attempted through the exclusion of the most variable positions. We present the gene order and complete nucleotide sequence of the mitochondrial genomes of the South American and the Australian lungfish. The, comparisons of all three living lungfish mitochondrial genomes allows for the first time the definition of mtDNA synapomorphies of the Dipnoi. The major aim of this study was to resolve the long-standing question about the identity of the closest living lineage among fish to tetrapods (reviewed in Meyer 1995; Meyer and Zardoya 2003; Zardoya and Meyer 1997b). The addition of two other lungfish lineages leads to a better taxon sampling (Graybeal 1998; Hendy and Penny 1989; but see Poe and Swofford 1999) and is expected to divide the long- branch of

the previously single published lungfish sequence (Zardoya and Meyer 1996a) and therefore to reduce the potential effect of long-branch attraction artifacts (Zardoya and Meyer 2001c).

Materials and Methods

Isolation, PCR, Cloning, and Sequencing Procedures

DNA was extracted from white muscle tissue of a single South American (*Lepidosiren paradoxa*) and Australian (*Neoceratodus forsteri*) lungfish as previously described (Zardoya et al. 1995). The isolated mtDNA was used as a template for amplification by PCR (standard PCR conditions with 35 cycles): denaturing at 94°C for 60 s, annealing at 42–47°C for 60 s, and extending at 72°C for 120 s were performed in 25-μl reactions containing 2.5 μl 10× buffer, 2.5 μl MgCl$_2$ (10 mM), 2 μl dNTP (2.5 mM each), 1 μl each primer (10 pmol/μl), template DNA, and 0.625 μl TAQ (1 U/μl) (Sigma–Aldrich Co.). Conserved primers allowed us to directly sequence parts of the mitochondrial genome of both lungfish (see also supplementary data, Table, A 1A and 1B). They were previously designed to sequence the mitochondrial genome of the coelacanth (*Latimeria chalumnae*) and were expected to successfully amplify mitochondrial DNA fragments in other related vertebrate species (Zardoya and Meyer 1997a). Based on these DNA- fragments additional species-specific primers were designed.

PCR amplifications of long fragments were achieved by using the expand long template PCR system of Boehringer Mannheim (PCR buffer 1). The following cycle conditions were used: 1 cycle of 2 min at 94°C; 10 cycles of 10 s at 94°C, 30 s at 58°C, and 6 min at 68°C; 20 cycles of 10 s at 94°C, 30 s at 58°C, and 6 min at 68°C, with an extension step that increased 20 s for each cycle; 1 cycle of 5 min at 68°C; and, finally, 4°C ad infinitum. The resulting 3.3- to 6.7-kb fragments were used as DNA templates for subsequent standard PCR amplifications of shorter fragments to facilitate subsequent sequencing.

Sequencing of the PCR products was performed with an automated DNA sequencer (ABI PRISM 377) using the standard protocol for the Big-Dye Deoxy Terminator cycle sequencing kit (PE Applied Biosystems Inc.). The sequences were generated for both strands from each clone, averaging 550 bp in length, and each sequence overlapped the next by about 50–100 bp. In no case sequence differences were observed between the overlapping regions.

Phylogenetic Analyses

For the phylogenetic analyses 30 complete mitochondrial genomes were selected to have about an equal number of taxa for all the major groups of vertebrates and their sequences were subdivided into three data sets: (1) all 12 mitochondrial proteins that are encoded on the H-strand (except ND6, which is encoded on the L-strand), (2) all 22 tRNA genes, and (3) the combined 12S and 16S rRNA genes. The alignments were first generated by ClustalX (Thompson et al. 1997) and subsequently refined by eye. In all three data sets all positions that contain gaps in more than two of the sequences were eliminated, leading to 1488 and 2394 nucleotide positions (starting with 1635 and 2942 positions) in the tRNA and the rRNA data sets, respectively.

The protein data set was analyzed at the amino acid level. In this case in addition to the elimination of positions containing more than two gaps all highly variable and therefore probably noisy positions were excluded from the phylogenetic analyses. This resulted in 2902 remaining amino acid positions starting from a total

of 3593 positions. All three data sets were analyzed by maximum parsimony (MP), minimal evolution (ME) with gamma-corrected distance estimates, and maximum likelihood (ML) analyses using the program PhyML (mtrev model), taking rate variations across sites (RAS) into account (Guindon and Gascuel 2003), and, finally, Bayesian inference (BI). For the protein data set only the MP analysis was done with PAUP* (Swofford 1993, 1999). The estimation of the gamma parameter for the maximum likelihood analyses was done for the protein data set with Tree-Puzzle version 5 (Strimmer and Von Haeseler 1996). The ME bootstrap analysis with gamma parameter-based distance estimate was performed with MEGA version 2.0 (Kumar 2001).

In addition, the separate tRNA, rRNA, and protein (aa-level) data were combined and analyzed with MrBayes (version 3.04b), which permitted a simultaneous analysis at both the DNA and the protein level. For this analysis the gamma-distributed rates, the proportions of invariant position, the stationary frequencies (for both amino acids and nucleotide), and the replacement and substitution frequencies were unlinked for the three partitions of the data set. We also performed an additional analyses with the combined data set, unlinking the branch lengths for the three partitions. These later analyses showed that the branch lengths differ considerable among the three data partitions (data not shown). The phylogenetic tree obtained in the analyses of the combined data set is very similar to the one obtained with the protein data set, therefore we assume that even the combined analysis is not containing a strong signal for any of the basal nodes (data not shown).

For the tRNA and rRNA data sets the MP and ME analyses were conducted with PAUP*, the ML bootstrap analyses were done with PhyML, and the Bayesian inference with MrBayes. MrBayes version 2.01 implemented on a G4 PowerPC was used for the Bayesian inference with 100,000 generations, trees were sampled every 10 generations (Huelsenbeck and Ronquist 2001). The number of generations needed until convergence around a stable likelihood value was in the range of 5–10% for the nucleotide data sets and usually higher for the protein data set with 10–15%. The reliability of the resulting topologies was estimated for all three data sets using bootstrap analyses (MP and ME, with 2000 replicates each; ML, only 100 replicates; MP, with 10 times random addition and TBR options), and Bayesian posterior probabilities given as percentages. In addition, the significance of alternative topologies, especially for the protein data set, was estimated by the Kishino–Hasegawa (1989) test and the Shimodaira–Hasegawa (Ota et al. 2000) Likelihood ratio tests, implemented in the Tree Puzzle software package version 5.1 (Schmidt et al. 2002).

Results and Discussion

Genome Organization

The complete sequences of the mitochondrial genomes from the Australian (AuL) and South American (SAL) lungfish have been deposited in GenBank under the accession numbers AF302933 for AuL and AF302934 for SAL. The relative position and orientation of the 2 rRNAs, 22 tRNAs, 13 protein coding regions, and the control region in both new lungfish sequences are found to be identical to the vertebrate consensus mitochondrial gene order (see supplementary data, Table 2A and 2B). The lengths of the mitochondrial genomes are 16,403 and 16,572 bp for the South –American and the Australian lungfish, respectively.

Ribosomal RNA Genes

The 12S and 16S rRNA genes in the South American lungfish mitochondrial genome are 929 and 1577 bp long—and differ in size by only 8 and 14 bp from the lengths of those genes in the African lungfish. In the Australian lungfish, the 12S and 16S rRNA genes are 952 and 1680 bp long—15 and 89 bp longer, respectively, than in the African lungfish. Comparisons of our data to previously published mitochondrial rRNA sequences (12S and 16S) with those from other individuals of the same species Lepidosiren paradoxa and Neoceratodus forsteri (Hedges 1993) show only minor sequence differences, of approximately 0.3 and 0.1%, respectively. The positions that were different from the previously published sequence from other lungfish individuals in the 12S and 16S rRNA genes were examined with special care. Differences were overall randomly distributed and only a single region in Neoceratodus at positions 1006–1011 is markedly different; here GTGCGC is found in our sequence and TCCG in the other.

Transfer RNA Genes

The typical set of 22 tRNA genes were found in the mitochondrial genomes of the South American and Australian lungfish. The tRNA genes range in size from 65 to 74 nucleotides, and all can potentially be folded into cloverleaf secondary structures permitting G–U pairings (see supplementary data, Fig. 2A and 2B). As previously described for the African lungfish (Zardoya and Meyer 1996a) and other tetrapods, e.g., the caecilian (Zardoya and Meyer 2000), high variability was found mainly in the DHU and TψC loops. As expected, particularly anticodon regions, but also acceptor stems were found to be more invariant. In contrast to the African and Australian lungfish and some other animals (Wolstenholme 1992), the tRNASer(AGY) of the South American lungfish forms a normal cloverleaf structure (although it exhibits only very weak pairings in the acceptor stem). Therefore, the unusual structure of a reduced DHU arm in tRNASer(AGY) does not appear to be suitable as a synapomorphy to describe the lungfish clade as was previously suggested (Kumazawa and Nishida 1993; Zardoya and Meyer 1996a).

Protein-Encoding Genes

All expected 13 large open reading frames were detected in the mitochondrial genome of both Lepidosiren paradoxa and Neoceratodus forsteri. Except for ND6, all proteins were encoded on the heavy

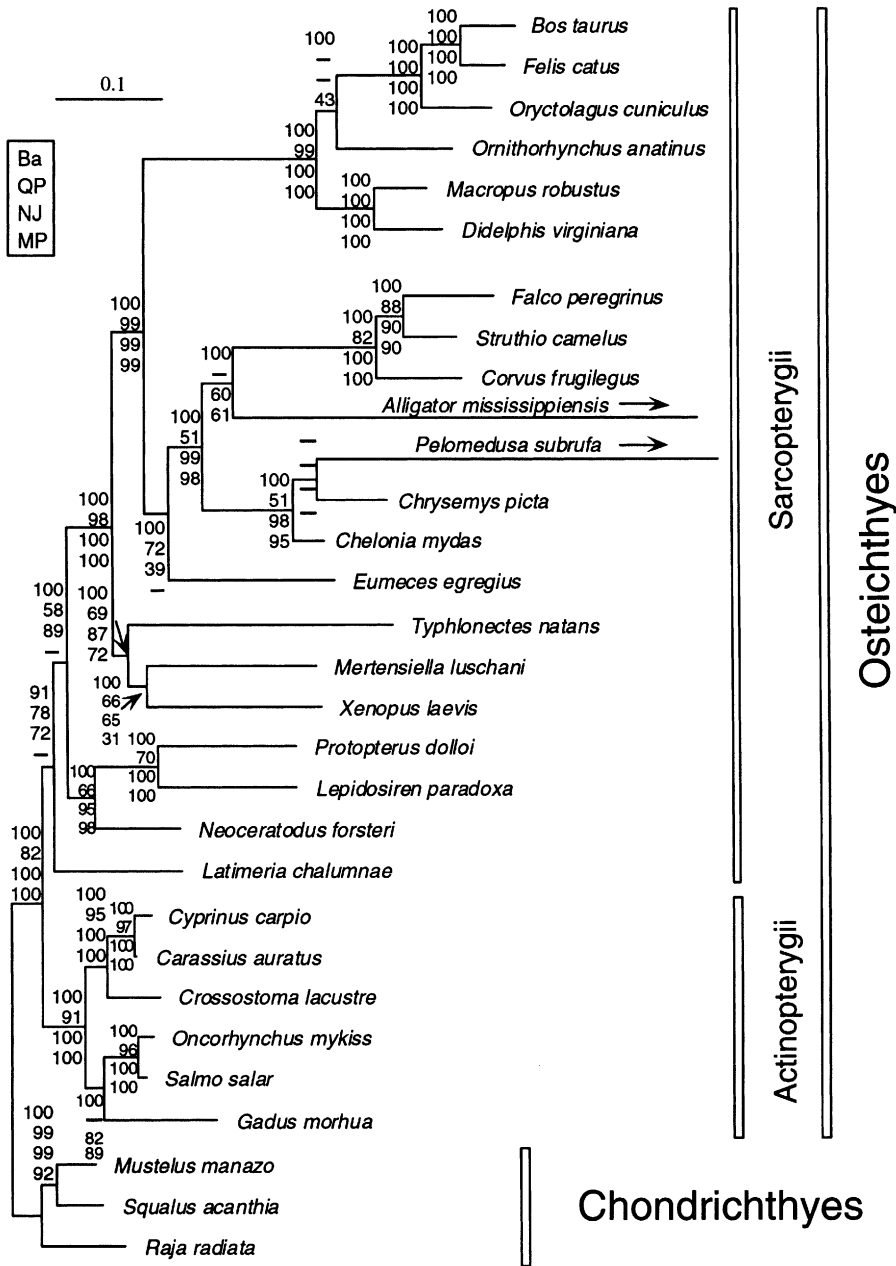# Total H-Strand Mitochondrial Proteins (12)

Fig. 2. Phylogenetic tree based on the deduced amino acid sequences of all mitochondrial protein coding genes (except ND6) showing the relationships between actino- and sarcopterygian fish including tetrapods. The branch lengths were estimated by the program Tree Puzzle under a MTREV + F + Γ + I model with eight rate categories. The data set was analyzed using four phylogenetic methods, BI, ML, NJ, and MP. The robustness of the groups was tested using 2000 bootstrap replicates each for NJ and MP (TBR; 10 times random addition) and 100 replicates for ML. The ML bootstrap estimates were obtained using the program PhyML (Guindon and Gascuel 2003). The Bayesian inference was conducted using MrBayes. The following applies to all shown phylogenetic trees: Chondreichthyes were used as the outgroup. Bootstrap values or BI poster probabilities (given as percentages) lower than 40% were either not shown at all or replaced by a dash, if at least one of the three other support values for this node is higher than 40%. Note that the internal nodes connecting the major groups are very short.

strand. Just like in the mitochondrial genome of the African lungfish (Zardoya and Meyer 1996a) and other vertebrates except for the lamprey (Lee and Kocher 1995), there are two cases of reading frame overlap in two genes encoded by the same strand (ATPases 8 and 6 overlap, 10 bp [*Lepidosiren, Protopterus*] and 9 bp (*Neoceratodus*); ND4L and ND4 overlap, 7 bp [*Lepidosireno Neoceratodus, Protopterus*]). In *Lepidosiren* and *Neoceratodus* as well as in *Protopterus*, ATG is found to be the start codon for all protein-encoding genes, except of CO1, which is initiated by GTG. Most of the genes in the mitochondrial genome of the South American lungfish end with a TAA stop codon (ND1, ND2, CO1, ATPase8, ATPase6, CO3, ND4L, ND5, Cyt b), three genes end with a TAG (ND3, ND4, ND6), and only one uses AGA as a stop codon (CO2). AGA has never been found to be used as a stop codon in ray-finned fish and could therefore be interpreted as a synapomorphy of the sarcopterygian lineage that deliniates it from the actinopterygians. In the Australian lungfish nearly the same stop codon usage of TAA (ND1, CO2, ATPase8, ATPase6, CO3, ND4L, ND5, Cyt b), TAG (ND2, CO1, ND3, ND6), and AGA (ND4) is observed as in the South American lungfish.

Table 2. List of the complete mitochondrial genomes used in this study

Scientific name	Common name	Accession no.	Reference
Felis catus	Cat	NC_001700	Lopez et al. (1996)
Bos taurus	Cow	NC_001567	Anderson et al. (1982)
Oryctolagus cuniculus	Rabbit	NC_001913	Gissi et al. (1998)
Macropus robustus	Wallaroo	NC_001794	Janke et al. (1997)
Didelphis virginiana	Opossum	NC_001610	Janke et al. (1994)
Ornithorhynchus anatinus	Platypus	NC_000891	Gemmell et al. (1994)
Struthio camelus	Ostrich	NC_001953	Harlid et al. (1997)
Corvus frugilegus	Rook	NC_002069	Harlid and Arnason (1999)
Falco peregrinus	Falcon	NC_000878	Mindell et al. (1999)
Alligator mississippiensis	Alligator	NC_001922	Janke and Arnason (1997)
Chelonia mydas	Turtle	NC_000886	Kumazawa and Nishida (1999)
Chrysemys picta	Turtle	NC_002073	Mindell et al. (1999)
Pelomedusa subrufa	Turtle	NC_001947	Zardoya and Meyer (1998)
Eumeces egregius	Skink	NC_000888	Kumazawa and Nishida (1999)
Xenopus laevis	Frog	NC_001573	Roe et al. (1985)
Mertensiella luschani	Salamander	NC_002756	Zardoya and Meyer (2001b)
Typhlonectes natans	Caecilian	NC_002471	Zardoya and Meyer (2000)
Protopterus dolloi	African lungfish	NC_001708	Zardoya and Meyer (1996a)
Latimeria chalumnae	Coelacanth	NC_001804	Zardoya and Meyer (1997a)
Cyprinus carpio	Carp	NC_001606	Chang et al. (1994)
Carassius auratus	Carp	NC_002079	Murakami et al. (1998)
Crossostoma lacustre	Loach	NC_001727	Tzeng et al. (1992)
Gadus morhua	Atlantic cod	NC_002081	Johansen and Bakke (1996)
Oncorhynchus mykiss	Trout	NC 001717	Zardoya et al. (1995)
Salmo salar	Salmon	NC_001960	Hurst et al. (1999)
Squalus acanthias	Spiny dogfish	NC_002012	Rasmussen and Arnason (1999)
Mustelus manazo	Dogfish	NC_000890	Cao et al. (1998)
Raja radiata	Skate	NC_000893	Rasmussen and Arnason (1999)

Why Is the Tetrapod Origin Question So Difficult to Solve?

Despite much previous work on the molecular phylogenetic relationships of the relevant groups, the answer to the important question, which of the two surviving groups of sarcopterygian fish, the lungfish or the coelacanth, is the closest living relative of the land vertebrates is still somewhat uncertain (reviewed in Meyer and Zardoya 2003). Previous molecular phylogenetic studies that used mitochondrial DNA data sets were able only to unambiguously rule out the traditional textbook hypothesis that the coelacanth is the living sister group to land sarcopterygians (reviewed in Meyer and Zardoya 2003; Zardoya and Meyer 1997b). However, which of the other two hypotheses (Fig. 1), the lungfish as a sister group to tetrapods or the lungfish + coelacanth as a sister group, is correct still remained to be unambiguously determined. Recent nuclear DNA data sets, particularly based on the RAG genes, were successful in supporting the lungfish–tetrapod sister group relationship and, lately, in significantly rejecting the two alternatives topologies (Fig. 1) (Brinkmann et al. 2004; Venkatesh et al. 1999, 2001; but see Kikugawa et al. 2004; Takezaki et al. 2004).

Several factors contribute to the difficulty in resolving the relationships among the three sarcop-

terygian lineages. First, all three relevant lineages arose about 400 mya ago, within a narrow time window of perhaps 10–20 mya; hence, both the lungfish and the coelacanth lineages are almost equally 400 mya old. Second, only a single or possibly two closely related species in the case of the coelacanth (Erdmann et al. 1999; Holder et al. 1999; Pouyaud et al. 1999) and a small number of species (six) in the case of the lungfish are extant and therefore available for molecular phylogenetic investigation. Both groups represent the last survivors of formerly diverse and highly successful groups (Cloutier and Ahlberg 1996; Forey 1998a). Apparently, the split between the Actinopterygian and the Sarcopterygian lineages occurred only slightly earlier than that.

Long unbroken branches can result in artificial groupings in phylogenetic analyses (Brinkmann and Philippe 1999; Felsenstein 1978; Philippe et al. 2000). Long branches are, in this case, due to the fact that almost all species of these originally diverse groups went extinct and only very few or even only one or two very closely related species (in the case of the coelacanth) from these ancient lineages survived. Moreover, differences in branch lengths reflect pronounced differences in the rates of evolution between lineages. Most of the tetrapod lineages including all mammals and birds also appear to have (possibly 5 to 10 times) higher evolutionary rates (based on the
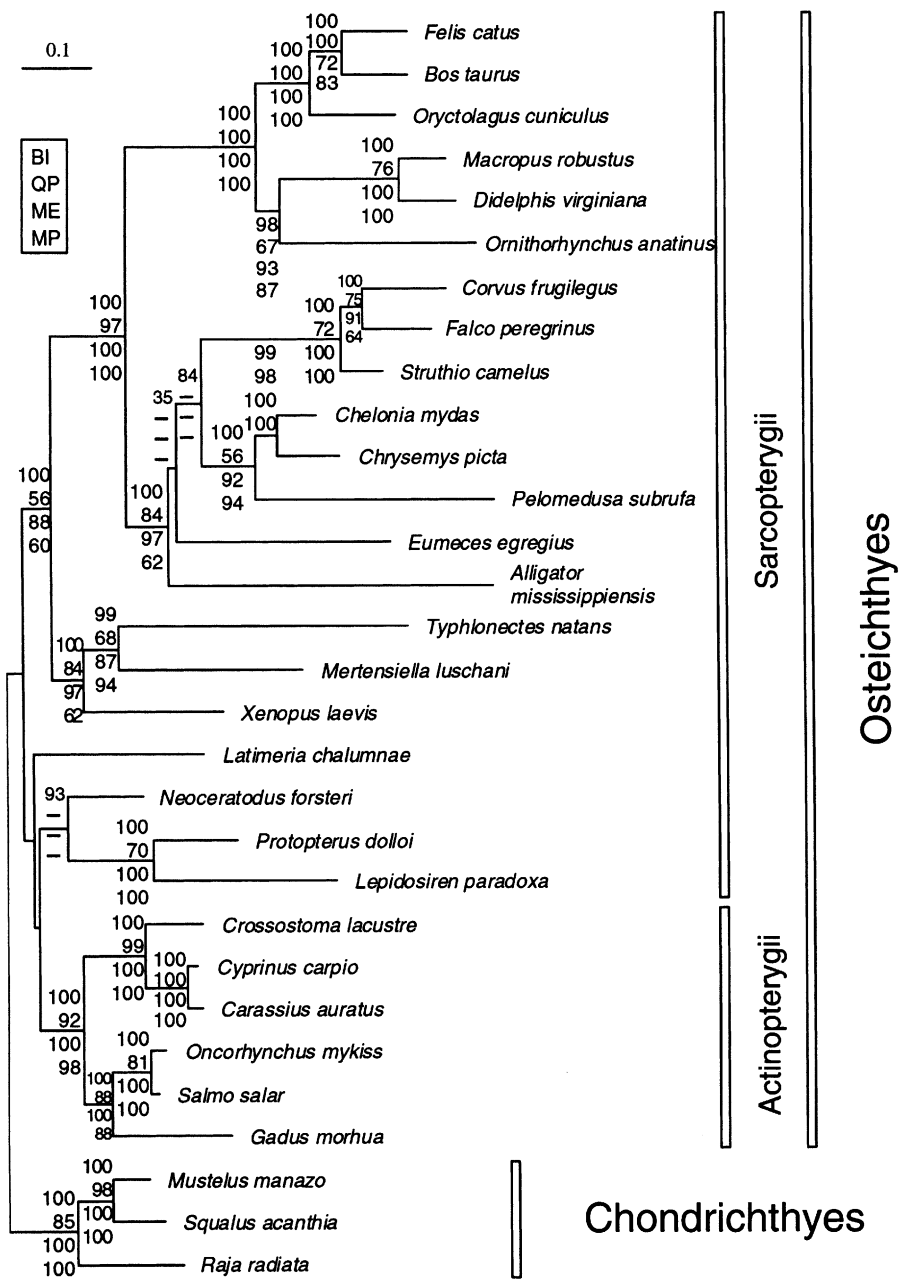
Total Mitochondrial tRNAs (22)

Fig. 3. Phylogenetic tree based on the same 30 mitochondrial genomes as in Fig. 2 based on all 22 mitochondrial tRNA genes. The branch lengths were estimated by the program PAUP* under a GTR + F + Γ + I model with eight rate categories. The phylogenetic methods used were identical except that nucleotide data were used. For the ME and MP analyses PAUP* was used, the ML bootstrap estimates were again done by the program PhyML, and the Bayesian inference was performed with MrBayes. Note the high bootstrap support of the caecilian + salamander sister-group relationship, the higher evolutionary rate of the caecilian sequence, and the absence of high bootstrap support for all deep branches connecting the major nontetrapod groups.

protein data set in Fig. 2) than the much more slowly evolving lineages of Neopterygii and Chondrichthyes (for nomenclature see Table 1).

To deal with these two difficult issues in phylogenetic tree reconstruction we increased the species sampling by adding the sequences of the remaining two lungfish lineages to our data set. The objective was to "break" the long lungfish branch and thereby to increase the confidence in the placement of lungfish in relation to the coelacanth and the tetrapod lineages. Of course, in the case of the long unbroken coelacanth branch we are unable to add taxa, since there is only one, and possibly a second, but too

closely related species of this once successful lineage alive. Furthermore, in the phylogenetic analyses we aimed to reduce the adverse effects of noise through stringent elimination of highly variable positions.

Phylogenetic Analyses Based on Complete Mitochondrial Genomes

The complete sequences of the mitochondrial genomes from the Australian and the South American lungfish allow a phylogenetic analysis based on a large number of positions and with a more complete taxon sampling than previous phylogenetic analyses

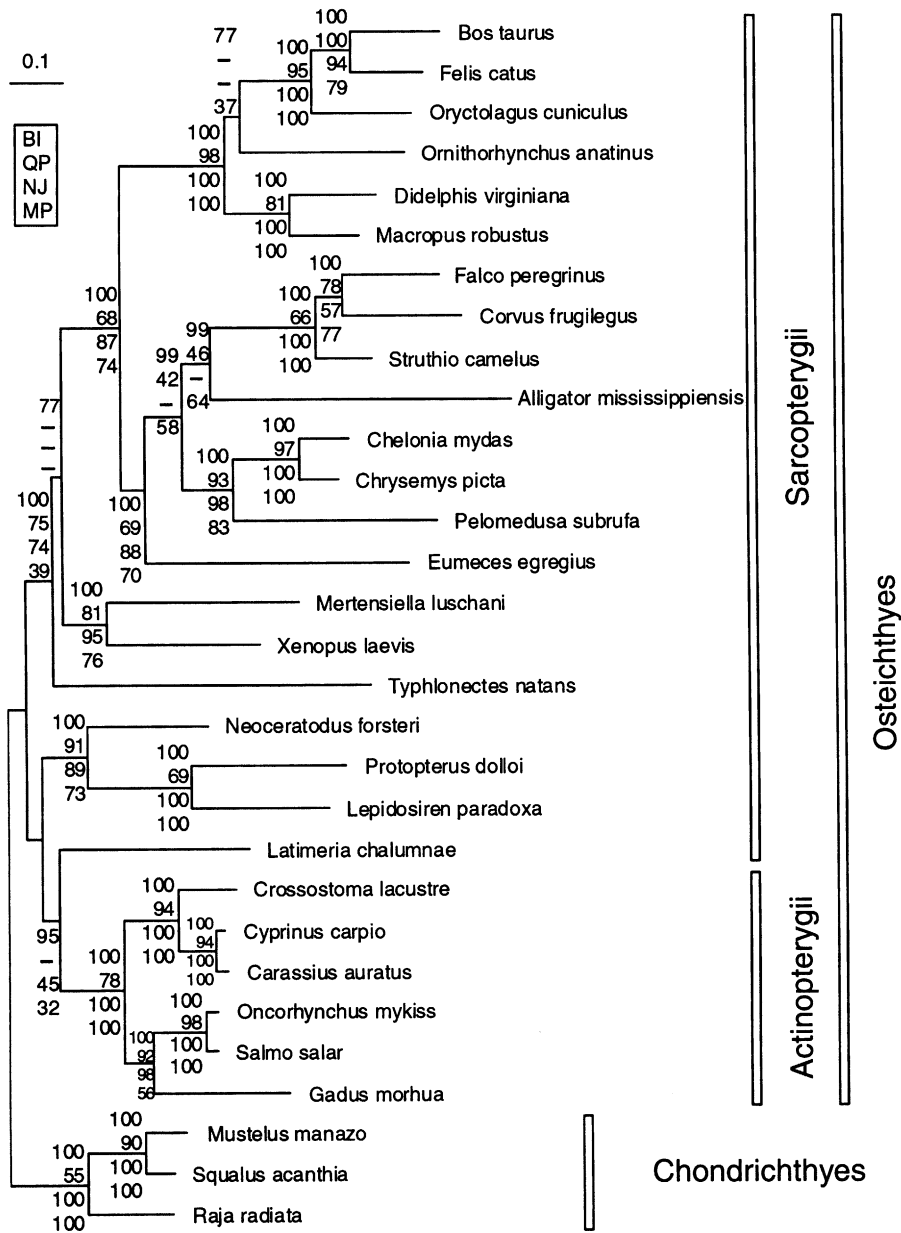## Total Mitochondrial 12S + 16S rRNAs

Fig. 4. Phylogenetic tree of the same species as in Figs. 2 and 3 based on the concatenated 12S and 16S mitochondrial rRNA genes. The branch lengths were estimated by the program PAUP* under a GTR + F + Γ + I model with eight rate categories. The methods used were identical to those described for Fig. 3. Note the nonmonophyly of the amphibians and the high artificial support for a specific salamander–frog sister-group relationship with the caecilian as a highly supported independent group (see further discussion).

on smaller sets of genes. Complete mitochondrial genome sequences are typically divided into three data sets (proteins, tRNAs, and rRNAs), which are composed of genes that are expected to have similar molecular evolutionary behavior within each set.

Our data set comprises 30 complete mitochondrial genome sequences and it is the first one that, not only includes the two new lineages of lungfishes, but also all three major amphibian groups including a caecilian (Zardoya and Meyer 2000) and a salamander (Zardoya et al. 2003; Zardoya and Meyer 2001c). The complete sequences of the mt genomes from 28 organisms obtained from GenBank (listed in Table 2) were aligned with the new South American and

Australian lungfish sequences from this study. Three cartilaginous fish mtDNA sequences were designated outgroups since their basal position in the vertebrates is not contended by phenotypic and molecular analyes (Kikugawa et al. 2004).

All three data sets were analyzed by gamma parameter-based maximum likelihood (ML) using PhyML as well as a Baysian inference (BI) approach, gamma parameter-based minimal evolution (ME) and maximum parsimony (MP). The topologies obtained by the different methods for the same data sets as well as for the different data sets were, generally, quite similar. However, some especially fast-evolving taxa (e.g., *Alligator*) could not be placed confidently

with any of the methods and in all data sets. For all three cases the ML tree with the highest likelihood value is shown and the support values obtained for this topology are indicated in the corresponding figures (see Figs. 2, 3, and 4).

As already suggested in previous studies (Corneli and Ward 2000; Cummings et al. 1995; Russo et al. 1996; Zardoya and Meyer 1996c), the combined protein data set (Fig. 2) seems to best resolve the phylogenetic relationships among vertebrates. It does so particularly well among the deep nodes and the protein data set typically yields, e.g., fully resolved topologies in the quartet puzzling analysis as implemented in Tree Puzzle (data not shown: 50% majority consensus tree). Accordingly, it is also the protein data set in which the most pronounced rate differences between taxa are observed (see, e.g., alligator versus birds and Chondrichthyes versus tetrapods) (Fig. 2). This observation might indicate that the protein data set is less saturated than the other two analyzed nucleotide data sets. In highly saturated molecular data sets usually all sequences have similar branch lengths, because all comparisons between basal sequences will lead to a comparable distance that is close to the value that random sequences would be expected to have. There is strong support for the monophyly of tetrapods in the protein data set, which is much less so the case in the two other nucleotide data sets, especially for the MP analyses (Figs. 3 and 4).

In general, the nucleotide data sets do not clearly resolve the "deepest" relationships between the major groups analyzed, this is among other things also reflected by the absence of strong support for these nodes and by conflicting support obtained with different reconstruction methods (Figs. 2 and 3). The support values for the deep nodes are mostly weak (<50%). Odd topologies, like, e.g., the lungfish or only the faster-evolving Lepidosireniformes as a sister group of the Chondrichthyes and the coelacanth as a sister group to the ray-finned fish, are, if found, at times supported by bootstrap values higher than 40%, yet the composition of these odd clades often varies based on which phylogenetic method was used.

In conclusion, the nucleotide data sets (rRNA and tRNA) are not suitable to resolve any of the deep-level relationships among the vertebrates. This observation is also confirmed by likelihood ratio tests, where all changes in the topology that affect the relative order of the major groups are far from significant (data not shown), indicating that the overall level of support for the topologies is rather weak. The nucleotide data sets provide more phylogenetic signal at an intermediate and more recent time scales, where they sometimes seem to perform even better than the protein data set. This is especially notable for the tRNA data set in the case of the tetrapod interrela-

tionships. Among tetrapods the rRNA data set clearly performs worse, because it is not recovering the monophyly of amphibians. In the tRNA data set, the lungfish monophyly is only recovered by the likelihood-based methods (ML, 30%; BI, 93%).

Amniote Phylogeny: Reptile Interrelationships

The amniotes as a monophyletic group are highly supported by all data sets; this also holds for the monophyly of mammals. The extant mammals radiated substantially later than all other major tetrapod lineages analyzed except the birds. The monophyly of reptiles is highly supported by both the tRNA and the rRNA data sets (see Figs. 3 and 4), but the support for this taxon based on the protein data is surprisingly weak in the ME and MP analyses (Fig. 2). The protein tree clearly favors the Archosauria hypothesis (birds and alligators form a monophyletic group). The support for this clade is less strong in the rRNA ML tree (not recovered by ME), whereas in the tRNA ML tree (Fig. 3) the situation is more complicated. The ML tree favors a bird–turtle sister-group relationship (as shown in Fig. 3), which is only supported by a moderate BI value of 84%, but all other support values slightly favor the Archosauria clade (ML, ME, and MP: 59, 66, and 62; data not shown) in agreement with the protein data (Fig. 2).

The protein tree (Fig. 2) suggests rather surprising turtle interrelationships; nevertheless, this topology is supported by a quite high delta likelihood value of 10.91 (SE, 12.05) against the topology with the expected sister-group relationship of *Chelonia* and *Chrysemys*. Additionally, there are no support values for the best topology shown in Fig. 2, whereas a sister-group relationship of *Chelonia* and *Chrysemys* is strongly supported by all phylogenetics methods used (BI, ML, ME, and MP: 100, 51, 100, and 100). This relationship is also highly supported in the tRNA and rRNA data sets (Figs. 3 and 4).

In the protein data set the support for the skink (as representative of the lepidosaurs) and not the turtle as the most basal lineage within the reptiles is very high (BI, ML, ME, and MP: 100, 99, 99, and 98). The basal position of the skink is also supported by high BI and ML values (99 and 93%) for the rRNA data and more weakly by MP (Fig. 4). For the tRNA data set, the alligator was determined to be even more basal than the skink in the ML tree, but the basal position of skink and alligator is only supported by BI (84%). In contrast to the tRNA best ML topology, the other bootstrap support values are weakly favoring a turtle + skink relationship (ML, ME, and MP: 55, 92, and 68) (Fig. 3). It is noteworthy that in none of the analyses a basal position of the turtles was observed, a position that is rather

commonly assumed to be correct (Zardoya et al. 1998; Zardoya and Meyer 2001a).

Amphibian Interrelationships

The protein data set strongly supports the monophyly of the tetrapods, irrespective of the phylogenetic method used. For the tRNA and the rRNA data sets (Figs. 3 and 4), the support is less pronounced but still rather strong, especially with the more sensitive likelihood-based methods. The protein and the tRNA data sets (Figs. 2 and 3) quite strongly support the monophyly of the amphibians for all methods used. However, the interrelationships among the three living amphibian orders remain somewhat unresolved. Either a sister-group relationship between frogs and salamanders, the Batrachia hypothesis, or a sister-group relationship between salamanders and caecilians is supported. The protein data set shows, except for the high BI value of 100%, low but consistent support in favor of the Batrachia hypothesis as is commonly assumed to be correct (reviewed in Meyer and Zardoya 2003). However, the tRNA data set supports quite strongly the sister-group relationship between salamanders and caecilians. This support is high and consistent (BI, ML, ME, and MP: 99, 92, 87, and 94%). The ML tree does not, in the rRNA data set, support the monophyly of amphibia but, instead, places the caecilian in a basal position, albeit with virtually no bootstrap support. Although there is a relatively high and consistent support for the Batrachia hypothesis, the rRNA data set is proposing the nonmonophyly of amphibian, albeit with virtual no support (Fig. 4).

The evolutionary rate of the caecilian is, in all three data sets and in all alternative topologies (data not shown), invariantly faster than those of the two other groups of amphibians. Therefore, a possible explanation for the conflicts between the different data sets may be a long-branch attraction (LBA) artifact, which interferes with the phylogenetic reconstruction. Under the assumption that the right topology is a sister-group relationship between Caudata and Gynophiona, one would expect that the LBA would ''pull'' the long branch, the caecilians, into a more basal position, a situation encountered only in the rRNA data set (Fig. 4). Under certain conditions, incorrect topologies may be sometimes recovered by all commonly used phylogenetic methods or may even be erroneously supported by high bootstrap values.

Two recent studies of amphibian relationships came to quite different results. Whereas Feller and Hedges (1998) found a highly supported sister-group relationship between caecilians and salamanders based on the 12S and 16S rRNA genes plus the tRNA^{VAL} and half of the tRNA$^{LEU(UUR)}$ (a total of about 2.7 kb), Zardoya and Meyer (2000) concluded that the present data (only 12S and 16S rRNA) does not allow resolution of the relationships among the three extant groups of amphibians conclusively. Possible reasons for the different results of the two studies have already been discussed (Zardoya and Meyer 2000) and are (i) the choice of the outgroup actinopterigian fish against amniotes and (ii) the presence of 11 (Zardoya and Meyer 2000) versus only 3 amniote species (Feller and Hedges 1998). The biggest difference to the rRNA analyses presented in this paper is that the two previous studies used three, instead of one, representative of the three main amphibian lineages. Our improved species sampling might have overcome the LBA and allowed us to recover the monophyly of amphibians with an rRNA data set. Zardoya and Meyer (2001b) based on the analyses of complete mitochondrial DNA genomes found support for the Batrachia hypothesis.

Sarcopterygian Relationships

The monophyly of the three lungfish lineages is highly supported by the protein and the rRNA data sets (Figs. 2 and 4), whereas in the case of the tRNA data set only the ML and the BI analyses support it (ML and BI, 93%; Fig. 3). In all three data sets we observe clearly higher evolutionary rates for the African and the South American lungfish compared to the Australian lungfish and the coelacanth, rendering the determination of the phylogenetic relationships among these taxa more difficult. The protein data set supports a lungfish + tetrapod sister group relationship (Figs. 1A and 2). Although this relationship is not indicated by overwhelmingly high bootstrap values, it is still the most strongly supported phylogenetic hypothesis for the three sarcopterygian lineages (Fig. 1). The support for the monophyly of the sarcopterygiian lineages is quite similar, although neither of the two nodes is supported by the MP analysis (Fig. 2).

There appears to be sufficient phylogenetic signal at least in the protein data set to resolve the sister-group relationship between sarcopterygians and actinopterygians, when cartilagenous fish are used as outgroup. The monophyly of the Actinopterygii is quite strongly supported (Figs. 2, 3, 4) in all of our analyses with all three mitochondrial data sets, and its sister-group relationship to the sacropterygia is well supported, at least in the protein-coding data set (Fig. 2). The monophyly of the actinopterygian and the sarcopterygian lineages is not found with any of the two nucleotide data sets, but this is not surprising because of their limited resolution concerning deep-level relationships (Figs. 3 and 4).
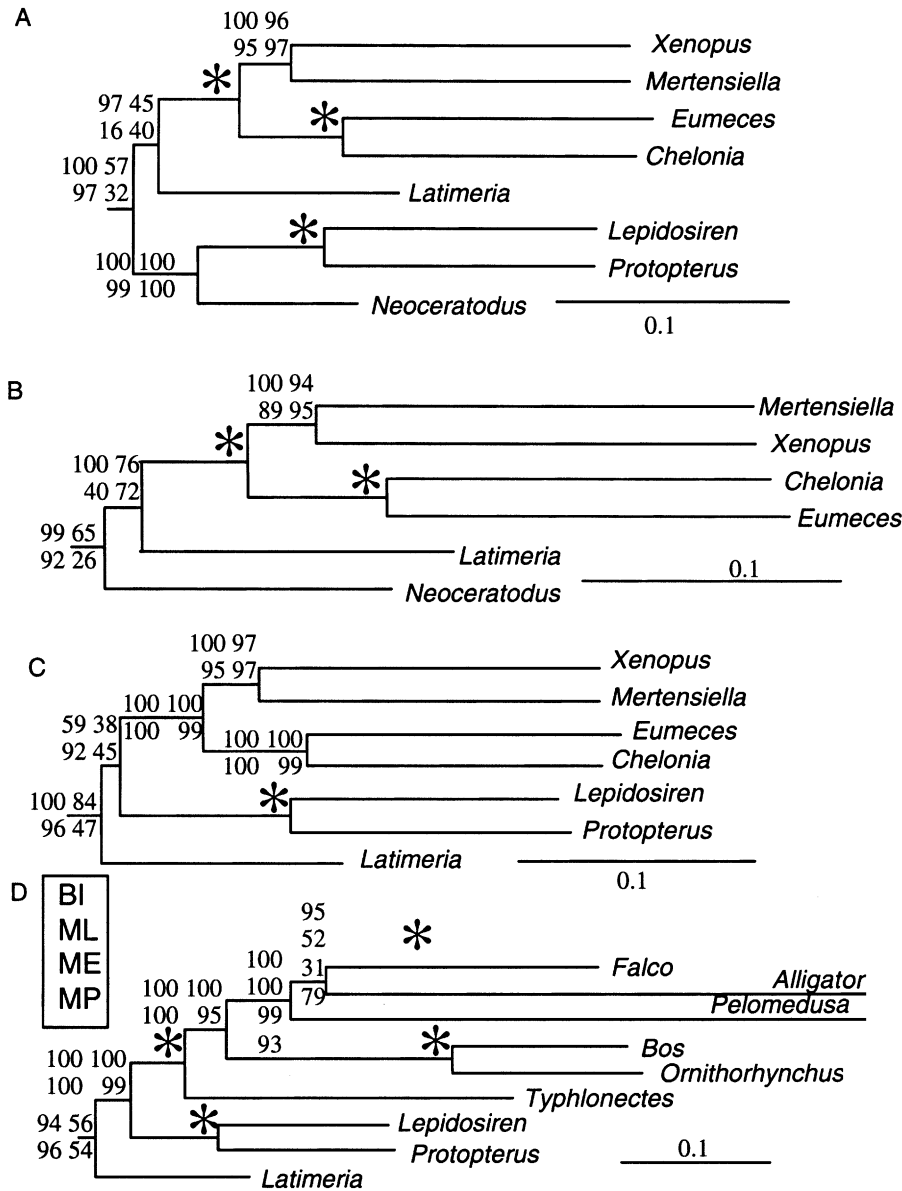
Fig. 5. Phylogenetic tree using the same number of amino acid positions (2902) as in Fig. 2, but with all fast-evolving tetrapod sequences (13 in total). Only the sarcopterygian part of the tree is shown in all cases. Asterisks indicate support values of 100% in all four analyses. The difference between **A** and **B** is that in B the sequences of the two fast-evolving lungfish are removed. In **C** the slowly evolving sequence of the Australian lungfish is eliminated, and in **D** the LBA of the tetrapod sequences is greatly enhanced by replacing the four slowest by the five fastest sequences.

Does LBA Influence the Position of the Lungfish?

The tetrapod sequences evolve considerably faster than the ones of all other groups analyzed. This fact is especially obvious for mammals and reptiles and can be best seen in the analyses of the protein data set (Fig. 2). The addition of the two newly established mitochondrial lungfish genome sequences successfully permitted a subdivision of the long and previously undivided branch of the African lungfish and, thereby, considerably reduced the possibility of LBA artifacts. Nevertheless, it became obvious that the mitochondrial sequences of both the African and the South American lungfish are roughly twice as fast evolving, than that of the Australian lungfish (see Fig. 2). This leads to the question whether the mere presence of the slowly evolving Australian lungfish is sufficient to exclude any LBA artifact between the two fast-evolving African and South American lungfish and the very fast tetrapod sequences. In order to address this question, we decided to reduce the possible LBA of the tetrapods to the presently accessible minimum, by the elimination of all fast-evolving (tetrapod) sequences (Fig. 5). Only the four most slowly evolving tetrapod sequences were included in these analyses, whereas all other groups were left unchanged. In all four parts of Fig. 5A–D only the sarcopterygian sequences of the trees are shown. Surprisingly, in Fig. 5A lungfish are no longer resolved as the closest relatives of the tetrapods but, rather, the coelacanth is. Although this topology is only weakly supported by bootstrap values, except for the BI (97%), it does nevertheless represent the best ML tree. This change in position of the lungfish in the tree (Fig. 5A) is suggestive of a LBA artifact because the elimination of the fast tetrapod sequences

Table 3. Likelihood ratio tests of the three alternative topologies (implemented in Tree Puzzle version 5.0): 1 = (sharks,(coelacanth,(lungfish,tetrapod))); 2 = (sharks,(lungfish,(coelacanth,tetrapod))); 3 = (sharks,((lungfish,coelacanth),tetrapod)))

Tree	logL	Difference	SE	p, 1sKH	p, SH
		Complete protein data set (30 taxa, 2902 aa positions)			
1	−46,612.96	0.00	← Best	1.0000 +	1.0000 +
2	−46,620.12	7.15	10.42	0.2320 +	0.4820 +
3	−46,622.43	9.46	10.06	0.16701 +	0.3760 +
		17 taxa without all fast tetrapods (2902 aa positions)			
1	26,472.20	4.77	9.27	0.3050 +	0.4230 +
2	26,467.43	0.00	← Best	1.0000 +	1.0000 +
3	26,474.99	7.55	8.12	0.1810 +	0.2870 +
		15 taxa without African and South American lungfish (2902 aa positions)			
1	−23,528.25	8.69	8.73	0.1530 +	0.2170 +
2	−23,519.57	0.00	← Best	1.0000 +	1.0000 +
3	−23,528.82	9.25	8.27	0.1420 +	0.1820 +
		16 taxa without *Neoceratodus* (2902 aa positions)			
1	−25,432.11	0.00	← Best	1.0000 +	1.0000 +
2	−25,432.80	0.69	7.00	0.4610 +	0.6480 +
3	−25,432.66	0.55	6.79	0.4620 +	0.6410 +
		18 species without *Neoceratodus* but with the 5 fastest-evolving tetrapods (2902 aa positions)			
1	−32,306.25	0.00	← Best	1.0000 +	1.0000 +
2	−32,336.66	30.41	11.02	**0.0030** −	**0.0040** −
3	−32,331.12	24.87	12.14	**0.0180** −	0.0210 −

Note. p, probability. p values <0.05 are considered significant and shown in boldface. aa, amino acid; KH, Kishino–Hasegawa test; SH, Shimodaira–Hasegawa likelihood ratio test.

considerably reduces the attraction of the remaining tetrapods. If, in addition, the two fast-evolving sequences of the African and the South American lungfish are removed as well (see Fig. 5B), the support for the coelacanth as a sister-group to tetrapods is further strengthened. In the two remaining analyses (Fig. 5C and D) we were trying the opposite by reinforcing the LBA, either by eliminating slowly evolving sequences or by replacing the slowest-evolving by the fastest-evolving tetrapod sequences. In Fig. 5C we show the effect of the exclusion of the sequence of the slowly evolving Australian lungfish, which should clearly strengthen possible LBA artifacts between the remaining lungfish and the tetrapods. As expected under LBA, the tetrapod + lungfish taxon is again recovered (Fig. 5C).

Likelihood Ratio Test Analyses

Table 3 shows the results of likelihood ratio tests for the three alternative topologies in Fig. 1 and for all five protein data sets analyzed (all 30 taxa; Fig. 5A, 17 taxa without fast tetrapods; Fig. 5B, 15 taxa without fast tetrapods + lungfish; Fig. 5C; 16 taxa without *Neoceratodus*; and Fig. 5D, 18 taxa without *Neoceratodus* but with 5 very fast tetrapods). Despite the varying support in favor of either a lungfish + tetrapod or a coelacanth + tetrapod sister-group relationship, in none of the first four data sets are any of the alternative topologies statistically significantly rejected (at a p value <0.05). Only in the last analysis

where we were trying to artificially maximize the effect of LBA by selecting the fastest-evolving tetrapod sequences shown in Fig. 5D was significant support for the lungfish as the closest relative of tetrapods, to the exclusion of the two alternative topologies, obtained (see also Table 3). The support is significant not only for the Kishino–Hasegawa test, but also for the much more conservative Shimodaira–Hasegawa likelihood ratio test. However, if the sequence of *Neoceratodus* is added to the data set in Fig. 5D, the support for any of the alternative solutions again becomes nonsignificant (data not shown). This result demonstrates the crucial importance of slowly evolving taxa in the attempt to overcome LBA artifacts. The significant support in the case of the data set in Fig. 5D can only be interpreted as an LBA artifact, obtained by a setup that maximizes the effects of LBA. These results highlight the limitations of the mitochondrial data sets for the tetrapod origin question. However, the absence of strong support for any solution in rigorous phylogenetic analyses (elimination of most highly saturated regions) might be interpreted as suggesting that there is no pronounced bias in the data set under realistic conditions (reasonable species sampling), which would lead to a highly supported but incorrect solution.

Ongoing phylogenetic work on nuclear genes that code for slowly evolving proteins might provide the long-sought answer to the sister-group relationship among the three living lineages of sarcopterygians the tetrapods, coelacanths, and lungfish (see, e.g., Brinkmann et al. 2004).

References

Ahlberg PE, Milner AR (1994) The origin and early diversification of tetrapods. Nature (London) 368:507–514

Ahlberg PE, Clack JA, Luksevics E (1996) Rapid braincase evolution between Panderichthys and the earliest tetrapods. Nature (London) 381:61–64

Anderson S, Bruijn Md, Coulson A, Eperon I, Sanger F, Young I (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717

Benton MJ (1990) Phylogeny of the major tetrapod groups: Morphological data and divergence dates. J Mol Evol 30:409–424

Bischoff TLWV (1840) *Lepidosiren paradoxa*: Anatomisch untersucht und beschrieben. Leipzig: Leopold Voss, P.R.1-34 Translation in Ana Sci Nature 14:116–159

Brinkmann H, Philippe H (1999) Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Mol Biol Evol 16:817–825

Brinkmann H, Venkatesh B, Brenner S, Meyer A (2004) Towards a solution of the tetrapod origin question: An approach based on concatenated nuclear-encoded genes. Proc Natl Acad Sci USA 101:4900–4905

Campbell NA (1987) Biology. Benjamin/Cummings, Menlo Park, CA

Cao Y, Waddell PJ, Okada N, Hasegawa M (1998) The complete mitochondrial DNA sequence of the shark Mustelus manazo: Evaluating rooting contradictions to living bony vertebrates. Mol Biol Evol 15:1637–1646

Carroll RL (1988) Vertebrate paleontology and evolution. Freeman, San Francisco

Chang Y, Huang F, Lo T (1994) The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. J Mol Evol 38:138–155

Cloutier R, Ahlberg P (1996) Morphology, characters and interrelationships of basal Sarcopterygians. In: Stiassny LRPaGDJ MLJ (eds) Interrelationships of fishes. Academic Press, San Diego, pp 445–479

Corneli PS, Ward RH (2000) Mitochondrial genes and mammalian pnylogenies: Increasing the reliability of branch length estimation. Mol Biol Evol 17:224–234

Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA sequence data in phylogenetic analysis. Mol Biol Evol 12:814–822

Curole JP, Kocher TD (1999) Mitogenomics: Digging deeper with complete mitochondrial genomes. Trends Ecol Evol 14:394–398

Erdmann MV (1999) An account of the first living coelacanth known to scientists from Indonesian waters. Environ Biol Fishes 54:439–443

Erdmann MV, Caldwell RL, Jewett SL, Tjakrawidjaja A (1999) The second recorded living coelacanth from north Sulawesi. Environ Biol Fishes 54:445–451

Feller A, Hedges S (1998) Molecular evidence for the early history of living amphibians. Mol Phylogenet Evol 9:509–516

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410

Forey P (1998a) History of coelacanth fishes. Chapman & Hall, London

Forey P (1998b) A home from home for coelacanths. Nature 395:319–320

Gemmell N, Janke A, Western P, Watson J, Paabo S, Graves J (1994) Cloning and characterization of the platypus mitochondrial genome. J Mol Evol 39:200–205

Gissi C, Gullberg A, Arnason U (1998) The complete mitochondrial DNA sequence of the rabbit, Oryctolagus cuniculus. Genomics 50:161–169

Graybeal A (1998) Is it better to add taxa or characters to a difficult phylogenetic problems? Syst Biol 47:9–17

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Harlid A, Arnason U (1999) Analyses of mitochondrial DNA nest ratite birds within the Neognathae: Supporting a neotenous origin of ratite morphological characters. Proc Roy Soc Lond Ser B Biol Sci 266:305–309

Harlid A, Janke A, Arnason U (1997) The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. Mol Biol Evol 14:754–761

Hedges SB, Hass CA, Maxson LR (1993) Relations of fish and tetrapods. Nature 363:501–502

Hendy M, Penny D (1989) A framework for the quantitative study of evolutionary trees. Syst Zool 38:297–309

Holder MT, Erdmann MV, Wilcox TP, Caldwell RL, Hillis DM (1999) Two living species of coelacanths?. Proc Natl Acad Sci USA 96:12616–12620

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Hurst C, Bartlett S, Davidson W, Bruce I (1999) The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. Gene 239:237–242

Janke A, Arnason U (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, Feldmaier-Fuchs G, Thomas W, von Haeseler A , Paabo S (1994) The marsupial mitochondrial genome and the evolution of placental mammals. Genetics 137:243–256

Janke A, Xu X, Arnason U (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, Bakke I (1996) The complete mitochondrial DNA sequence of Atlantic cod (*Gadus morhua*): Relevance to taxonomic studies among codfishes. Mol Mar Biol Biotechnol 5:203–214

Kikugawa K, Katoh K, Kuraku S, Sakurai H, Ishida O, Iwabe N, Miyata T (2004) Basal jawed vertebrate phylogeny inferred from multiple nuclea DNA-coded genes. BMC Biol 2:3

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

Kumar STK, Jakobsen IB, Nei M (2001) MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J Mol Evol 37:380–398

Kumazawa Y, Nishida M (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: Statistical evidence for Archosaurian affinity of turtles. Mol Biol Evol 16:784–792

Lee W-J, Kocher TD (1995) Complete sequence of a sea lamprey (Petromyzon marinus) mitochondrial genome: Early establishment of the vertebrate genome organization. Genetics 139:873–887

Lopez JV, Cevario S, O'Brien SJ (1996) Complete nucleotide sequences of the domestic cat (Felis catus) mitochondrial genome

and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. Genomics 33:229–246

Marshall CR (1987) Lungfish pylogeny and parsimony. In: Bemis WE, Burggren WW, Kemp NE (eds) The biology and evolution of lungfishes. Alan R. Liss, New York, pp 151–162

Meyer A (1995) Molecular evidence on the origin of tetrapods and the relationships of the coelacanth. Trends Ecol Evol 10:111–116

Meyer A, Dolven SI (1992) Molecules, fossils, and the origin of tetrapods. J Mol Evol 35:102–113

Meyer A, Wilson AC (1990) Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J Mol Evol 31:359–364

Meyer A, Zardoya R (2003) Recent advances in the (molecular) phylogeny of vertebrates. Ann Rev Ecol Syst 34:311–338

Mindell D, Sorenson M, Dimcheff D, Hasegawa M, Ast J, Yuri T (1999) Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst Biol 48:138–152

Murakami M, Yamashita Y, Fujitani H (1998) The complete sequence of mitochondrial genome from a gynogenetic triploid "Ginbuna" (Carassius auratus langsdorfi). Zool Sci (Tokyo) 15:335–337

Ota R, Waddell PJ, Hasegawa M, Shimodaira H, Kishino H (2000) Appropriate likelihood ratio tests and marginal distributions for evolutionary tree models with constraints on parameters. Mol Biol Evol 17:798–803

Panchen AL (1987) Character diagnosis of fossils and the origin of tetrapods. Biol Rev Phil Soc 62:341–438

Patterson C (1980) Origin of tetrapods: Historical introduction to the problem. In: Panchen A (ed) The Terrestrial environment and the origin of Land Vertebrates. Academic Press, New York

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Muller M, Le Guyader H (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc R Soc Lond B Biol Sci 267:1213–1221

Poe S, Swofford DL (1999) Taxon sampling revisited. Nature (London) 398:299–300

Pouyaud L, Wirjoatmodjo S, Rachmatika I, Tjakrawidjaja A, Hadiaty R, Hadie W (1999) A new species of coelacanth. Genetic and morphologic proof. CR Acad Sci III 322:261–267

Rasmussen A-S, Arnason U (1999) Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. J Mol Evol 48:118–123

Roe B, Ma D, Wilson R, Wong J (1985) The complete nucleotide sequence of the Xenopus laevis mitochondrial genome. J Biol Chem 260:9759–9774

Romer AS (1966) Vertebrate paleontology. University of Chicago Press, Chicago

Rosen DE, Forey PL, Gardiner BG, Patterson C (1981) Lungfishes, tetrapods, paleontology, and plesiomorphy. Bull Am Nat Hist Mus 167:159–276

Russo CAM, Takezaki N, Nei M (1996) Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol Biol Evol 13:525–536

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Smith JLB (1939a) A living fish of mesozoic type. Nature 143:455–456

Smith SLB (1939b) The living coelacanthid fish from South Africa. Nature 143:748–750

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

Swofford DL (1993) PAUP: A computer program for phylogenetic inference using maximum parsimony. J Gen Physiol 102:9A

Swofford DL (1999) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA

Takezaki N, Figueroa F, Zaleska-Rutcynska Z, Takahata N, Klein J (2004) The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of forty-four nuclear genes. Mol Evol Biol 21:1512–1524

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Tzeng C, Hui C, Shen S, Huang P (1992) The complete nucleotide sequence of the Crossostoma lacustre mitochondrial genome: Conservation and variations among vertebrates. Nucleic Acids Res 20:4853–4858

Venkatesh B, Ning Y, Brenner S (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. Proc Natl Acad Sci USA 96:10267–10271

Venkatesh B, Erdmann MV, Brenner S (2001) Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. Proc Natl Acad Sci USA 98:11382–11387

Wolstenhoime DR (1992) Animal mitochondrial DNA: Structure and evolution. Int Rev Cytol 141:173–216

Zardoya R, Meyer A (1996a) 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

Zardoya R, Meyer A (1996b) Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. Proc Natl Acad Sci USA 93:5449–5454

Zardoya R, Meyer A (1996c) Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol Biol Evol 13:933–942

Zardoya R, Meyer A (1997a) The complete DNA sequence of the mitochondrial genome of a "living fossil," the coelacanth (Latimeria chalumnae). Genetics 146:995–1010

Zardoya R, Meyer A (1997b) Molecular phylogenetic information on the identity of the closest living relative(s) of land vertebrates. Naturwissenschaften 84:389–397

Zardoya R, Meyer A (1998) Complete mitochondrial genome suggests diapsid affinities of turtles. Proc Natl Acad Sci USA 95:14226–14231

Zardoya R, Meyer A (2000) Mitochondrial evidence on the phylogenetic position of caecilians (Amphibia: Gymnophonia). Genetics 155:765–775

Zardoya R, Meyer A (2001a) The evolutionary position of turtles revised. Naturwissenschaften 88:193–200

Zardoya R, Meyer A (2001b) On the origin of and phylogenetic relationships among living amphibians. Proc Natl Acad Sci USA 98:7380–7383

Zardoya R, Meyer A (2001c) Vertebrate phylogeny: Limits of inference of mitochondrial genome and nuclear rDNA sequence data due to an adverse phylogenetic signal/noise ratio. In: Ahlberg PE (ed) Major events in early vertebrate evolution: Paleontology, phylogeny, and development. Taylor & Francis, London, pp 135–155

Zardoya R, Garrldo-Pretierra A, Bautista JM (1995) The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, Oncorhynchus mykiss. J Mol Evol 41:942–951

Zardoya R, Cao Y, Hasegawa M, Meyer A (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, Malaga-Trillo E, Veith M, Meyer A (2003) Complete nucleotide sequence of the mitochondrial genome of a salamander, Mertensiella luschani. Gene 317:17–27

Zimmer C (1999) At the waters edge: Fish with fingers, whales with legs Simon and Schuster, New York