The Complete Nucleotide Sequence of the Mitochondrial Genome of the Lungfish (*Protopterus dolloi*) Supports Its Phylogenetic Position as a Close Relative of Land Vertebrates

Rafael Zardoya and Axel Meyer

Department of Ecology and Evolution and Program in Genetics, State University of New York, Stony Brook, New York 11794-5245

Manuscript received June 10, 1995
Accepted for publication January 13, 1996

ABSTRACT

The complete DNA sequence (16,646 bp) of the mitochondrial genome of the African lungfish, *Protopterus dolloi,* was determined. The evolutionary position of lungfish as possibly the closest living relative among fish of land vertebrates made its mitochondrial DNA sequence particularly interesting. Its mitochondrial gene order conforms to the consensus vertebrate gene order. Several sequence motifs and secondary structures likely involved in the regulation of the initiation of replication and transcription of the mitochondrial genome are conserved in the lungfish and are more similar to those of land vertebrates than those of ray-finned fish. A novel feature discovered is that the putative origin of L-strand replication partially overlaps the adjacent $tRNA^{Cys}$. The phylogenetic analyses of genes coding for tRNAs and proteins confirm the intermediate phylogenetic position of lungfish between ray-finned fishes and tetrapods. The complete nucleotide sequence of the African lungfish mitochondrial genome was used to estimate which mitochondrial genes are most appropriate to elucidate deep branch phylogenies. Only a combined set of either protein or tRNA mitochondrial genes (but not each gene alone) is able to confidently recover the expected phylogeny among vertebrates that have diverged up to but not over ~400 mya.

THE transition from life in water to life on land, ~360 mya (BENTON 1990), was one of the most consequential events in the history of vertebrates. It was accompanied by a variety of refined morphological and physiological modifications, *e.g.,* reductions and rearrangements of the skull bones and modifications of swimming fins into load-bearing limbs (*e.g.,* PANCHEN and SMITHSON 1987). Two groups of lobe-finned fish (Table 1), the lungfish and the coelacanth (*Latimeria chalumnae*) have both been implicated as the closest living relative of tetrapods (reviewed in MEYER 1995). Lungfish were discovered >150 years ago (BISCHOFF 1840) and for several reasons, *e.g.,* they are obligate airbreathers, were initially believed to be amphibians, not fish. In the lower Devonian (~400 mya) lungfish were a species-rich group that inhabited both marine and freshwater environments (*e.g.,* reviewed in CLOUTIER 1991). However, only a very small number of "relict" species survive today. These are the Australian lungfish, *Neoceratodus forsteri,* the South American lungfish, *Lepidosiren paradoxa,* and four species in the genus *Protopterus* from Africa. These living fossils are of interest to evolutionary biology since their morphology, physiology, and biochemistry might be representative of that of the common ancestor of all land vertebrates. Therefore, lungfish have been widely studied by paleontologists, comparative morphologists and recently developmental biologists.

The other extant group of lobe-finned fish, the coelacanths were believed to have gone extinct ~80 mya, but in 1938, the only surviving species of this lineage of fishes was discovered off the coast of East Africa. Since its sensational rediscovery, the coelacanth is often depicted in textbooks as the "missing link" between fish and all land vertebrates *i.e.,* amphibians, reptiles, birds and mammals (ROMER 1966). However, many morphological, paleontological (*e.g.,* reviewed in PATTERSON 1980; ROSEN *et al.* 1981), and most molecular (MEYER and WILSON 1990; MEYER and DOLVEN 1992; HEDGES *et al.* 1993; reviewed in MEYER 1995; but see YOKOBORI *et al.* 1994) data suggest that lungfish and not the coelacanth are more closely related to tetrapods. Hence, the nucleotide sequence of the lungfish mitochondrial genome is of interest, both in terms of the evolution of the mitochondrial gene order in vertebrates and in terms of the phylogeny of land vertebrates.

Until now, the complete mitochondrial DNA sequences of 19 vertebrate species have been reported. Thirteen of them are from mammals, four from fishes, but only one of an amphibian and one of a bird have been determined. Remarkably, the structure and organization of vertebrate mitochondrial genomes is quite conserved and only minor rearrangements have been described for the chicken (DESJARDINS and MORAIS

Corresponding author: Rafael Zardoya, Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245. E-mail: rzardoya@life.bio.sunysb.edu

TABLE 1

Systematic position of lungfish

Class: Osteichthyes (bony fish)
Subclass: Actinopterygii (ray-finned fish)
Chondrostei (sturgeon, Acipenser; bichir, Polypterus)
Neopterygii (gar, Lepisosteus; bowfish, Amia; modern ray-finned fish, Teleostei)
Subclass: Sarcopterygii (lobe-finned fish)
Actinistia (coelacanth, Latimeria)
Rhipidistia
Dipnoi (lungfish)
Porolepiformes[a]
Osteolepiformes[a]
Tetrapoda (land vertebrates)
Lissamphibia (modern amphibians)
Amniota (reptiles, birds, mammals)

Modified from CARROLL (1988) and AHLBERG (1991). Position of lungfish in bold.

[a] Extinct.

1990) and the opossum (JANKE *et al.* 1994). Changes in gene order seem to be associated with the potential capability of tRNAs to translocate, although sometimes other genes are involved in transpositions as well. It is not yet clear when the establishment of the vertebrate consensus gene order occurred during their evolution. The lamprey, one of the earliest vertebrates, has a peculiar gene order (LEE and KOCHER 1995). Although it is similar to that of other vertebrates, it has enough differences (*i.e.,* two rather than one major noncoding regions, the missing O_L in the WANCY region, and translocations of cytochrome *b,* and $tRNA^{Pro}$, $tRNA^{Thr}$, and $tRNA^{Glu}$ are found) to be considered to be a unique and possibly a derived condition from the vertebrate consensus mitochondrial gene order. Despite the slow rate of evolution of gene order, nucleotide sequence evolution of animal mitochondrial DNA is rapid. The dynamic evolution of mitochondrial DNA sequences occurs through the accumulation of point mutations and make them particularly valuable for estimating phylogenetic relationships among closely related species (BROWN *et al.* 1979). However, not all mitochondrial genes evolve at the same rate (*e.g.,* reviewed in MEYER 1993), and some genes are more appropriate than others for inferring phylogenetic relationships among distantly related species.

We determined the complete nucleotide sequence and the gene order of the African lungfish mitochondrial genome. The aims of this study were to reconstruct mitochondrial genome evolution, estimate which mitochondrial genes (tRNA, rRNA, or protein-coding) are most appropriate to elucidate deep branch phylogenies, and clarify lungfish relationships to tetrapods and to ray-finned fish (Actinopterygii, Table 1). We are presently sequencing the coelacanth mitochondrial genome, with the long-term objective of establishing whether the lungfish or the coelacanth is the closest living relative of land vertebrates.

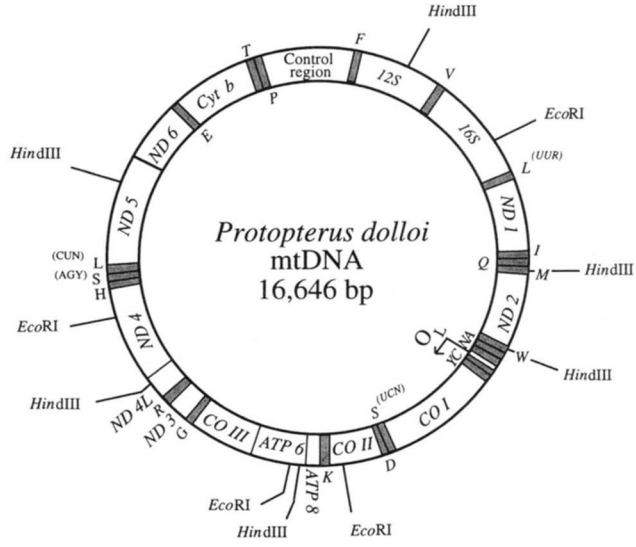

FIGURE 1.—Restriction map and gene organization of the *Protopterus dolloi* mitochondrial genome. All protein coding genes are encoded by the H-strand with the exception of *ND6,* which is coded by the L-strand. Each tRNA gene is identified by the single letter amino acid code and depicted according to the coding strand. Only the *Eco*RI and *Hin*dIII restriction sites used for cloning are shown.

MATERIALS AND METHODS

Mitochondrial DNA was purified from fresh eggs of a single individual of the African lungfish (*Protopterus dolloi*) as previously described (ZARDOYA *et al.* 1995a). After homogenization, intact nuclei and cellular debris were removed by a low-speed centrifugation (1000 × *g*). Mitochondria were pelleted by spinning at 10,000 × *g* for 20 min and subjected to a standard alkaline lysis procedure followed by a phenol/chloroform extraction. The isolated mtDNA was cleaved with *Eco*RI and *Hin*dIII restriction enzymes (see Figure 1 for positions of restriction sites). Three *Eco*RI fragments of 7.4, 3.1 and 0.7 kb and five *Hin*dIII fragments of 3.3, 2.9, 2.9, 2.2 and 1.1 kb were cloned into pUC18 covering the entire lungfish mtDNA molecule. In addition, the 7.4-kb *Eco*RI fragment was subcloned with *Sau*3A to facilitate sequencing.

Plasmid DNA was extracted from each clone using a Magic miniprep kit (Promega). After ethanol precipitation, cloned DNA was used as template for Taq Dye Deoxy Terminator cycle-sequencing reactions (Applied Biosystems Inc.) following manufacturer's instructions. Sequencing was performed with an automated DNA sequencer (Applied Biosystems 373A *Stretch*). Sequences were obtained using both M13 universal sequencing primers and 40 specific oligonucleotide primers. The sequences obtained from each clone were ~350 bp in length and each sequence overlapped the next contig by ~100 bp. In no case were differences in sequence observed between the overlapping regions. The location and sequence of these primers will be provided by the authors upon request.

Sequence data were analyzed by use of the GCG program package (DEVEREUX *et al.* 1984) and alignments and phylogenetic analyses were performed using CLUSTAL V (HIGGINS and SHARP 1989), PAUP Version 3.1.1 (SWOFFORD 1993), PHYLIP Version 3.5 (FELSENSTEIN 1989), and MOLPHY Version 2.2 (ADACHI and HASEWAGA, 1992).

RESULTS AND DISCUSSION

Genome organization: The complete sequence of the L-strand of the lungfish mtDNA is shown in Figure

2. The total length of the mitochondrial molecule is 16,646 bp. The overall base composition of the L-strand is A: 29%; T: 29%; C: 26%; and G: 16%. As in other vertebrates, two rRNAs, 22 tRNAs and 13 proteins are encoded by the lungfish mitochondrial genome; the relative position and orientation of all genes and the control region are identical to the vertebrate consensus mitochondrial gene order (Figures 1 and 2, Table 2). Peptide-encoding genes were identified by comparison with rainbow trout mtDNA (ZARDOYA *et al.* 1995a) and by the presence of initiation and stop codons. Sequences encoding tRNA genes were recognized by their capability to fold into putative cloverleaf structures, the presence of specific anticodons and by comparison with the rainbow trout homologues.

Noncoding sequences: The control region in the lungfish mitochondrial genome is 1184 bp long and it is localized between the *tRNA*^{Pro} and *tRNA*^{Phe} genes (Figure 2). In other vertebrates, this region usually includes the origin of H-strand replication and the sites of initiation of both H- and L-strand transcription. Analysis of the control region sequence permitted the identification of two conserved sequence blocks (CSB-II and -III) in the right domain, by comparison to the motifs reported by WALBERG and CLAYTON (1981) (Figure 2). A total of three termination associated sequences (TASs) were postulated in the left domain based on the consensus sequence proposed by DODA *et al.* (1981) (Figure 2). A putative CSB-I can be tentatively identified at position 16,247, but as in frog (ROE *et al.* 1985), this motif is reduced to only five nucleotides (GACAT) and shares limited sequence similarity to the human and mouse consensus sequence (WALBERG and CLAYTON 1981). The lungfish mitochondrial control region is also characterized by the presence of three 25-bp repeats at the 5′ end. These repeats are separated by TASs and their sequences are nearly identical to those of mammalian control regions (Table 3), but are less conserved or not found at all in frog, chicken, and ray-finned fish. These sequences contain the conserved motif 5′-TACAT-3′ and its complementary 5′-ATGTA-3′, which are proposed to maintain secondary structures in other control regions (SACCONE *et al.* 1991). Moreover, these pentanucleotide motifs seem to be associated with the presence of repeats in the left domain of control regions as seen in bats (WILKINSON and CHAPMAN 1991), shrews (STEWART and BAKER 1994), and sheep (ZARDOYA *et al.* 1995b) repeats. We identified a 40-bp sequence in the central domain (starting at position 16,013), close to the B block as defined by SOUTHERN *et al.* (1988) that is characterized by a 75–80% similarity with the corresponding sequences of sheep (ZARDOYA *et al.* 1995b), cow (ANDERSON *et al.* 1982), two species of seals (ARNASON and JOHNSSON 1992; ARNASON *et al.* 1993), rhinoceros (JAMA *et al.* 1993), pig (MACKAY *et al.* 1986), dolphin (SOUTHERN *et al.* 1988) and several species of whales (ARNASON *et al.* 1991; ARNASON and GULLBERG 1993; DILLON and WRIGHT 1993), but not found in other vertebrates.

The putative origin of light strand replication (O_L) is located in a cluster of five tRNA genes (WANCY region) (Figures 1 and 2) and is 45 nucleotides long. This region has the potential to fold into a stem-loop secondary structure with a stem formed by 15 paired and two unpaired nucleotides and a loop of 13 nucleotides. Half of the O_L stem is part of the *tRNACys* gene (Figure 3). Since this tRNA is encoded by the L-strand it seems likely that the same sequence is involved both in replicative and transcriptional events. This condition is not found in fish or amphibians (SEUTIN *et al.* 1994) suggesting that it might be a special feature of lungfish. The lungfish O_L loop contains a C-T rich sequence. This suggests that the initiation of L-strand synthesis is probably initiated in a polypyrimidine tract as in other fish (*e.g.*, JOHANSEN *et al.* 1990; ZARDOYA *et al.* 1995a), rather than and not restricted to a stretch of thymines as had been previously suggested for mammals (WONG and CLAYTON 1985).

Ribosomal RNA genes: The *12S* and *16S rRNA* genes in lungfish mitochondria are 937 and 1591 nucleotides long, respectively. Our sequence shows only minor differences to that previously reported for an unidentified species of *Protopterus* (HEDGES *et al.* 1993). More extensive divergence was observed relative to other species of the genus (*P. annectens* and *P. aethiopicus*) for portions of the *12S rRNA* gene that had previously been sequenced (MEYER and DOLVEN 1992). The primary sequence of both rRNA genes is alignable to that of other chordates (HEDGES *et al.* 1993) and the secondary structure appears to be conserved.

Transfer RNA genes: As in other vertebrates, the lungfish mitochondrial genome contains 22 tRNA genes interspersed between ribosomal RNA and protein coding regions. All the lungfish tRNA gene sequences can be folded into a cloverleaf secondary structure provided the formation of G-U wobble and other unusual pairings is allowed. These tRNAs range in size from 67 to 75 nucleotides, show high variability especially in their DHU and TψC loops, and are more constrained in their anticodon and acceptor stems. As in other animals, tRNA$^{Ser(AGY)}$ has a reduced DHU arm (WOLSTENHOLME 1992). On the other hand, tRNA$^{Ser(UCN)}$ and tRNA^{Lys} form a normal cloverleaf structure *e.g.*, in other fish, chicken and frog, strengthening the idea that the unusual structures inferred for these tRNAs in mammals can be considered synapomorphies that define this clade (KUMAZAWA and NISHIDA 1993). The proposed tRNA^{Cys} cloverleaf structure (Figure 3) indicates that this tRNA has a longer acceptor stem (8 bp instead of the usual 5 bp) and a shorter DHU stem (3 bp instead of the usual 4 bp) compared with any other vertebrate tRNA^{Cys}. Additional cloverleaf structures can be inferred yielding atypical DHU and TψC stems. If the *tRNA*^{Cys} gene acts as stem of the O_L, then constraints

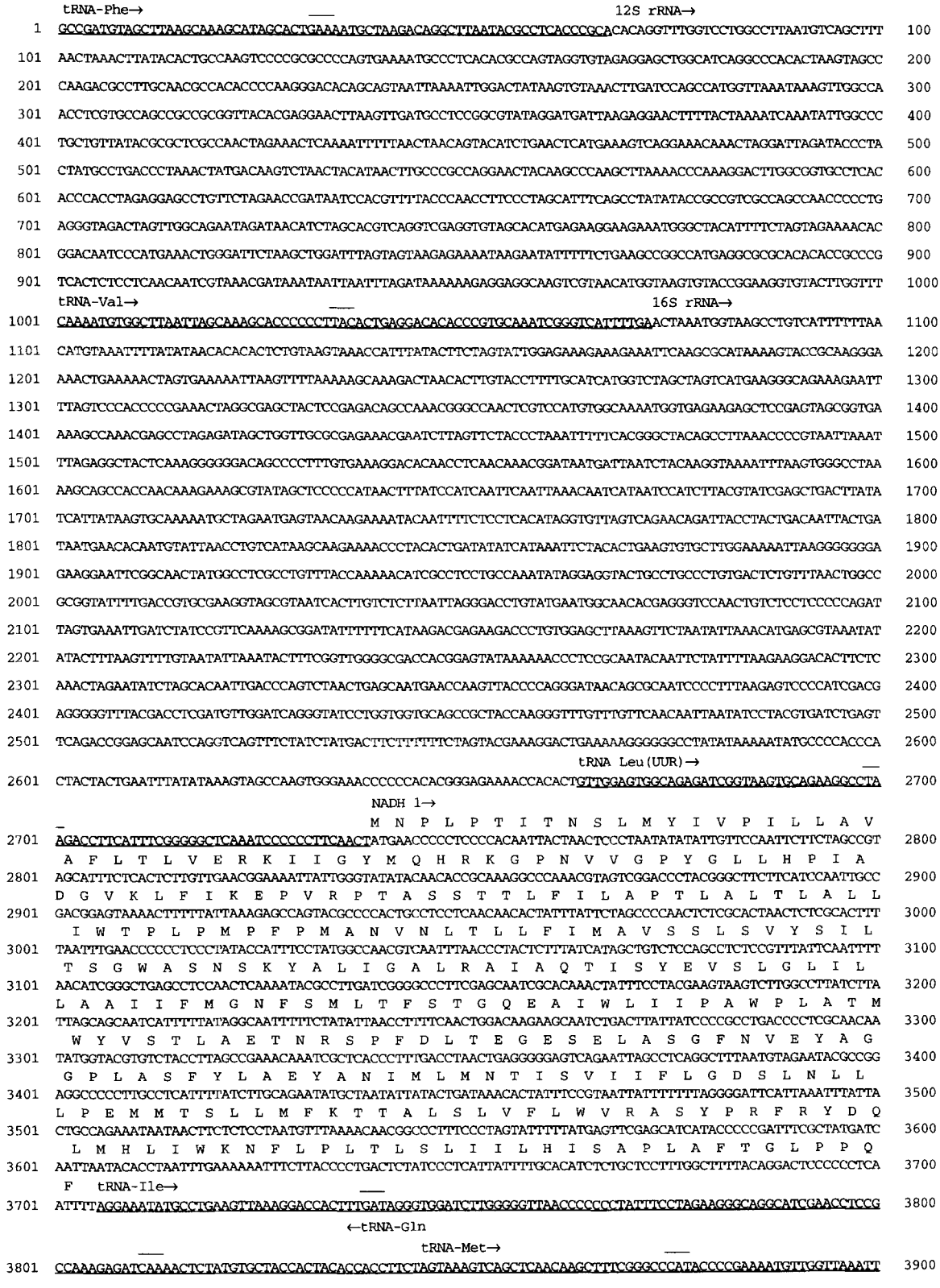

FIGURE 2.—Complete nucleotide sequence of the L-strand of the lungfish mitochondrial DNA. Position 1 corresponds to the first nucleotide of *tRNA*^{Phe}. Direction of transcription for each gene is denoted by arrows. The deduced amino acid sequence for each gene product is shown above the nucleotide sequence (one-letter amino acid abbreviation is placed above the first nucleotide of each codon). Termination codons are indicated by an asterisk. tRNA genes are underlined and the corresponding anticodons are overlined. In the control region CSB (Conserved Sequence Block) and TAS (Termination Associated Sequence) are underlined.

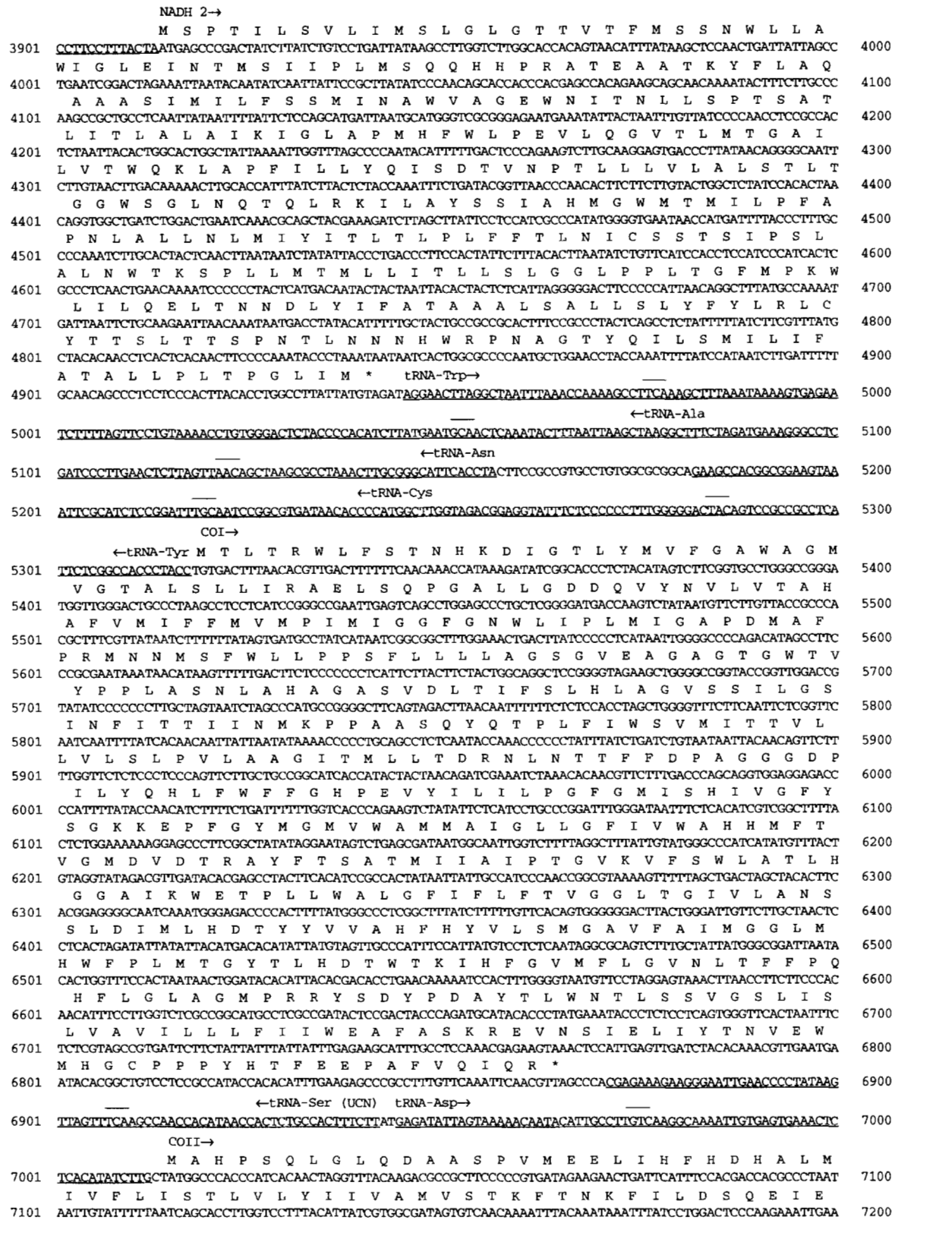

FIGURE 2.— *Continued*

will be added to the 5′ end of this gene leading to an unusual secondary structure of its product. As previously demonstrated (STEINBERG and CERDERGREN 1994), noncanonical structures such as that of $tRNA^{Cys}$ can be maintained by structural compensation within the tRNA molecule. An unusual cloverleaf for $tRNA^{Cys}$ has also been proposed in the reptile, *Sphenodon punctatus* (SEUTIN *et al.* 1994).

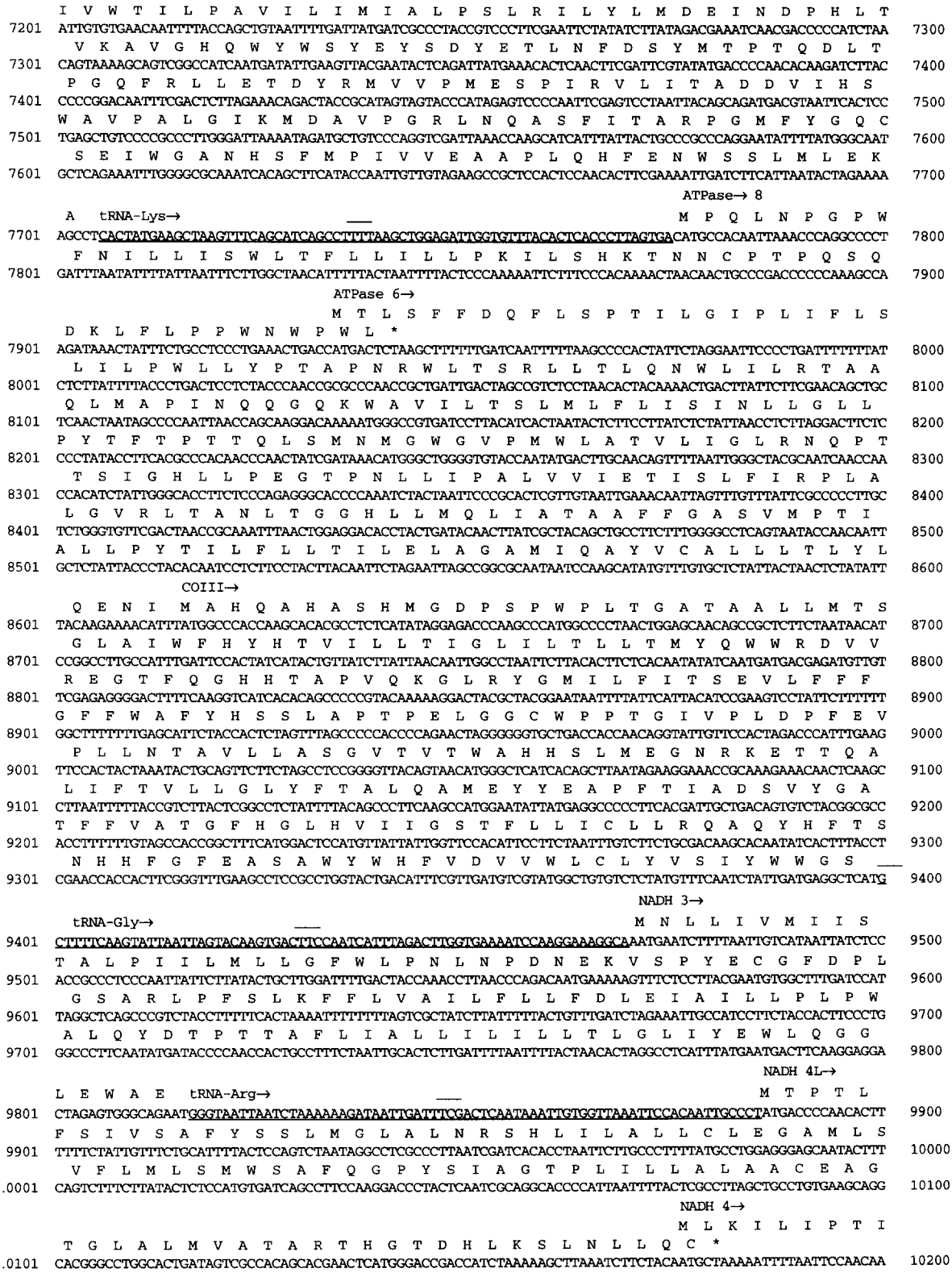

FIGURE 2.—*Continued*

Protein-encoding genes: The lungfish mitochondrial genome contains 13 large open reading frames (Figures 1 and 2) and, as in other vertebrate mtDNAs except 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 by 10 nucleotides; *ND4L* and *ND4* share seven nucleotides). All initiation codons in lungfish mtDNA protein-encoding genes are ATG except that of the *COI* gene, which is GTG (Table 2). This initiation codon usage is also shared by the

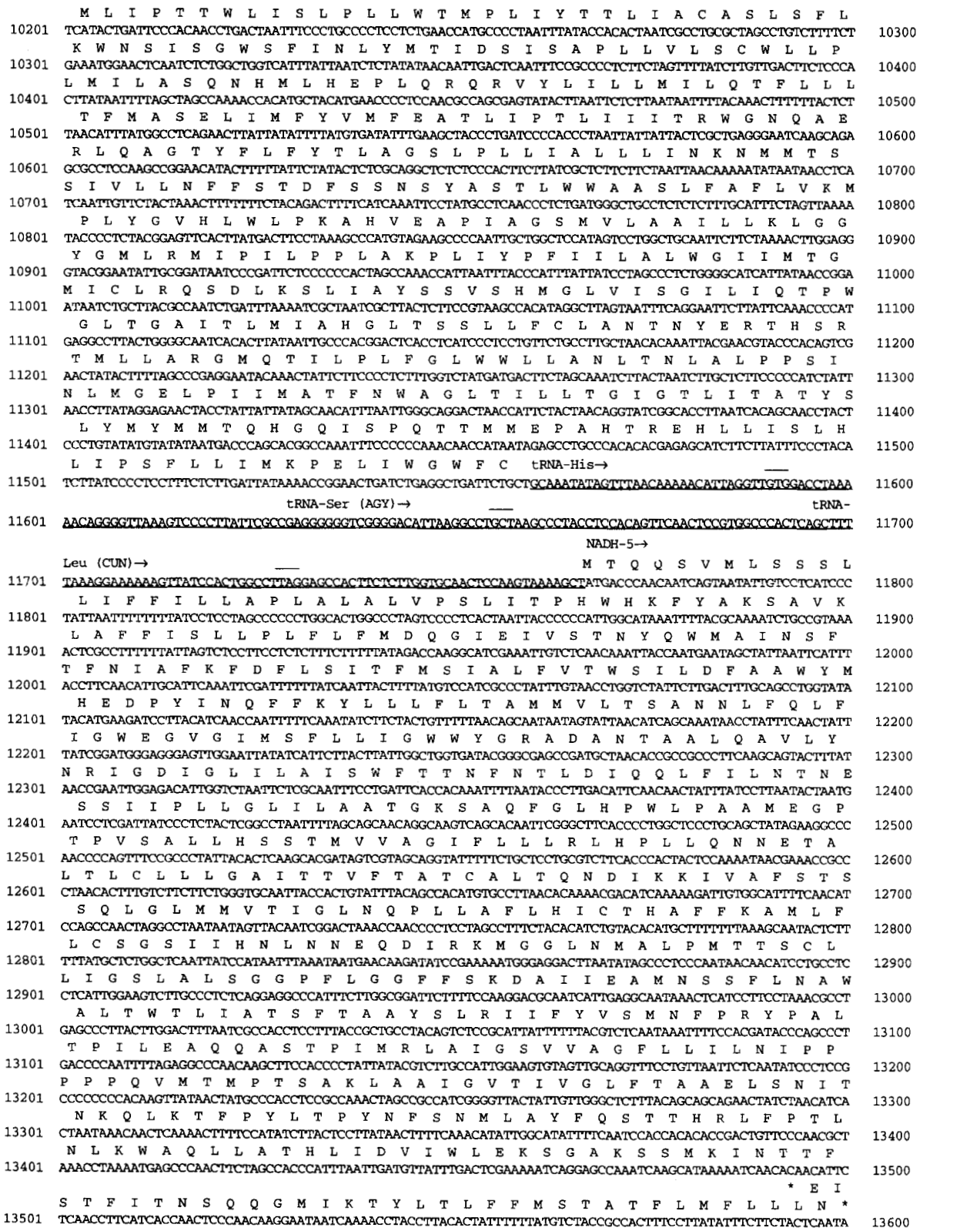

FIGURE 2.—*Continued*

four other fish mitochondrial genomes that have been completely sequenced (TZENG *et al.* 1992; CHANG *et al.* 1994; LEE and KOCHER 1995; ZARDOYA *et al.* 1995a) and by the chicken mitochondrial DNA (DESJARDINS and MORAIS 1990). Interestingly, most ORFs have "T" incomplete stop codons (*ND1, COII, ATPase 6, COIII, ND3, ND4* and *cyt b*), two end with TAA (*ATPase 8, ND4L*), three use TAG as stop codon (*ND2, COI* and *ND6*) and

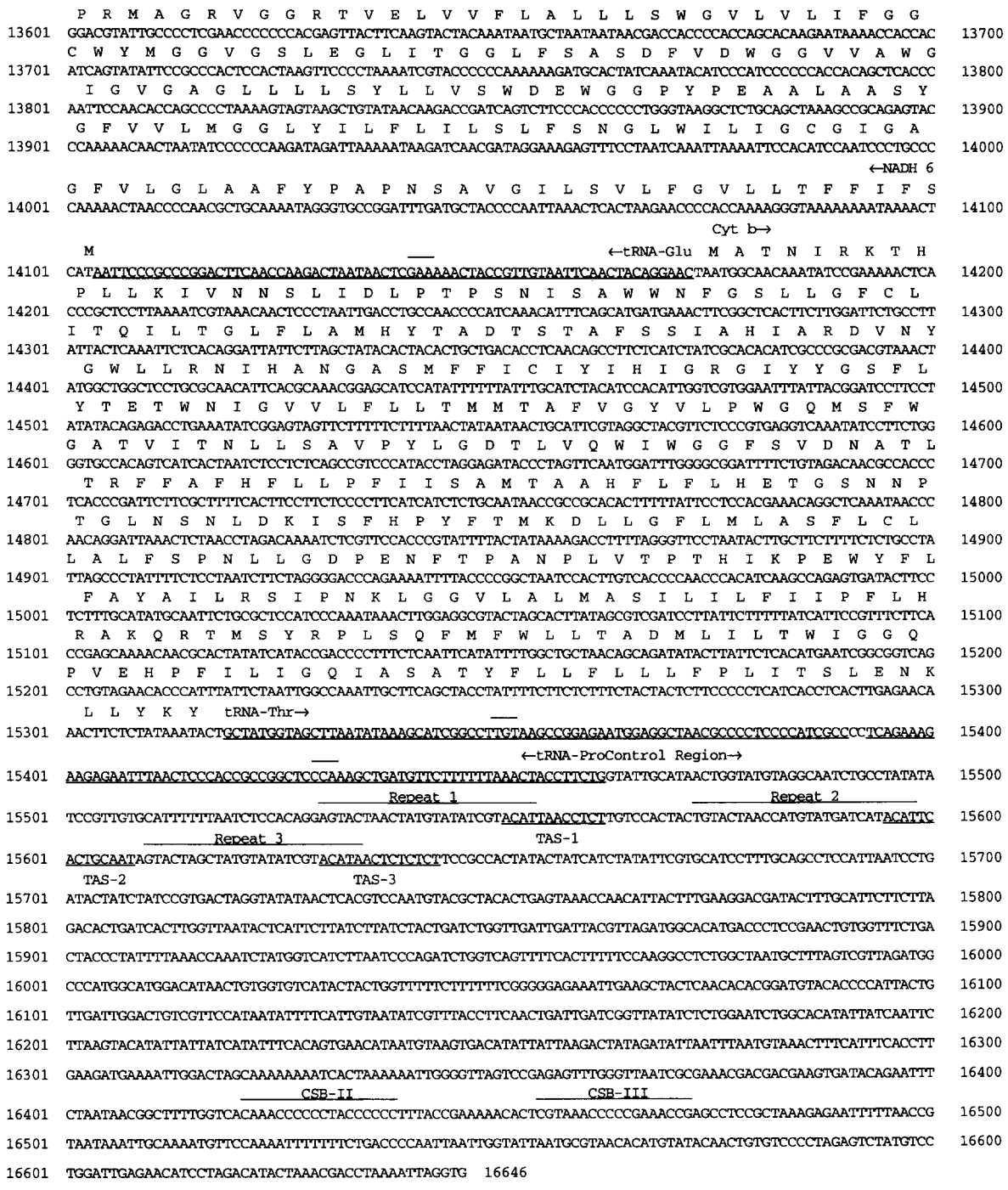

FIGURE 2.— *Continued*

one ends with AGG (*ND5*) (Table 2). So far, no ray-finned fish has been found that uses AGR as stop codon, whereas in frog (ROE *et al.* 1985) *ND5* ends with AGA.

The codon usage of the lungfish is similar to that of cod (JOHANSEN *et al.* 1990), loach (TZENG *et al.* 1992), carp (CHANG *et al.* 1994), lamprey (LEE and KOCHER 1995), rainbow trout (ZARDOYA *et al.* 1995a) and frog (ROE *et al.* 1985) (data not shown). As in other vertebrates (for review, see MEYER 1993), there is an evident bias against guanidine at the third codon position whereas there is an even distribution of the other three bases (Table 4). This anti-G bias is not as pronounced as in mammals and the lamprey and is similar to that of other fish and frog. In lungfish protein-coding genes, as in other mitochondrial genomes (Table 4), pyrimidines (%C + T = 68.0 ± 0.3) are overrepresented compared with purines in second codon positions. This pyrimidine bias in the second position directly reflects a hydrophobic bias in amino acid composition of mitochondrial proteins (NAYLOR *et al.* 1995) since most of the amino acid residues coded by NYN codons are hydrophobic. The hydrophobic bias of mitochondrial pro-

TABLE 2

Localization of features in the mitochondrial genome of the African lungfish

Feature	From	To	Size (bp)	Codon	
				Start	Stop
tRNA-Phe	1	67	67		
12S rRNA	68	1000	933		
tRNA-Val	1001	1072	72		
16S rRNA	1073	2663	1591		
tRNA-Leu (UUR)	2664	2738	75		
NADH 1	2739	3705	966	ATG	T--
tRNA-Ile	3706	3777	72		
tRNa-Gln	3845	3775	71 (L)		
tRNA-Met	3845	3913	69		
NADH 2	3914	4942	1028	ATG	TAG
tRNa-Trp	4945	5014	69		
tRNA-Ala	5083	5015	69 (L)		
tRNA-Asn	5156	5084	73 (L)		
tRNA-Cys	5248	5182	67 (L)		
tRNA-Tyr	5317	5249	69 (L)		
COI	5319	6866	1548	GTG	TAG
tRNA-Ser (UCN)	6941	6871	71 (L)		
tRNA-Asp	6944	7012	69		
CO II	7015	7705	691	ATG	T--
tRNA-Lys	7706	7774	69		
ATPase 8	7776	7943	168	ATG	TAA
ATPase 6	7934	8615	682	ATG	T--
CO III	8616	9399	784	ATG	T--
tRNA-Gly	9400	9469	70		
NADH 3	9471	9816	346	ATG	T--
tRNA-Arg	9817	9885	69		
NADH 4L	9886	10182	297	ATG	TAA
NADH 4	10176	11559	1384	ATG	T--
tRNA-His	11560	11628	69		
tRNA-Ser (AGY)	11629	11697	69		
tRNA-Leu (CUN)	11698	11766	70		
NADH 5	11767	13602	1836	ATG	AGG
NADH 6	14103	13591	513 (L)	ATG	TAG
tRNA-Glu	14172	14104	69 (L)		
Cyt*b*	14175	15318	1144	ATG	T--
tRNA-Thr	15319	15390	72		
tRNA-Pro	15462	15393	68 (L)		
Control region	15463	16646	1184		

Gene nomenclature according to ATTARDI *et al.* (1986). L, light-strand sense.

teins is due to their function as membrane-bound proteins involved in the electron transport chain (*e.g.*, ATTARDI *et al.* 1986).

Phylogenetic analyses of lungfish relationships: To correctly place lungfish among vertebrates, especially their relationship to ray-finned fish and tetrapods, the complete nucleotide sequences of the human (ANDERSON *et al.* 1981), blue whale (ARNASON and GULLBERG 1993), opossum (JANKE *et al.* 1994), chicken (DESJARDINS and MORAIS 1990), frog (ROE *et al.* 1985), carp (CHANG *et al.* 1994), loach (TZENG *et al.* 1992), trout (ZARDOYA *et al.* 1995a) and lamprey (LEE and KOCHER 1995) mitochondrial genomes were compared with that reported here. Protein-encoding genes were aligned and gaps were introduced according to the deduced amino acid sequences. Variation among the 13 protein coding genes was mainly found in the carboxyl-end of the polypeptides and in few cases in the amino-end. However, the central core of the mitochondrial proteins was found to be highly conserved. Therefore, ambiguous alignments at 5'- and 3'-ends of protein-coding genes were excluded from the phylogenetic analyses. Similarly, tRNA genes were aligned taking their secondary structures into account. In this case, DHU and TψC arms were omitted due to ambiguity in alignments; hence our reanalysis of KUMAZAWA and NISHIDA's data (1993) is not directly comparable with theirs. In all analyses, gaps in alignment were treated as missing data.

Tree reconstruction: Three different types of data sets were used to reconstruct phylogenetic trees. (1) A

TABLE 3

Conserved motif in the 5′ end of the control region

Sequence	Accession No.	Species
CTATGT-AT-ATCGTACATTAA	L42813	*Protopterus dolloi* (lungfish)
		Mammals
.....T.....	J01415	*Homo sapiens*
.....A.....	U12368	*Aepyceros melampus* (impala)
A.....A.....	J01394	*Bos taurus* (cow)
A.....AA.....	L29055	*Ovis aries* (sheep)
.....C.....G.....	D23665	*Equus caballus* (horse)
.....C.G....G.....	U03575	*Canis familiaris* (dog)
.....G.....	S68248	*Mirounga leonina* (elephant seal)
.....G.....	X63726	*Phoca vitulina* (seal)
.....G.....	L27310	*Taxidea taxus* (skunk)
.-.....T.....A...	X72204	*Balaenoptera musculus* (blue whale)
.....C.	L06553	*Sorex cinereus* (shrew)
T.....	X14848	*Rattus norvegicus* (rat)
T.....	U21162	*Mus musculus* (mouse)
.....G.....	X75874	*Ursus arctos* (bear)
.....A.....GG.	Z29573	*Didelphis virginiana* (oppossum)
		Reptiles
.....G...A..	U19540	*Sternotherus minor* (musk turtle)
.....G.....C.	L28795	*Graptemys pulchra* (map turtle)
T..A..C.....-.	U22261	*Caretta caretta* (loggerhead turtle)
		Amphibians
T.....ATAA.....	M57480	*Rana castebiana* (frog)
.....A...-.AG...A..	M10217	*Xenopus laevis* (clawed frog)
		Fish
.....T.A...CC.....	X54348	*Acipenser transmontanus* (sturgeon)
.....TA.CAC....C.	L07753	*Cyprinella spiloptera* (minnow)
A.....T...-A.....	M97985	*Salmo trutta* (trout)
A.....TA.CCC.....	U06060	*Jordanella floridae* (flagfish)
G.....TATCC.....	U06583	*Xiphophorus variatus* (swordtail)
A.....ATCAC.....	X17660	*Gadus morhua* (cod)
A.....AATCAC.....	U12069	*Pollachius virens* (pollock)

The repeats found in the 5′ end of the control region of the mitochondrial genome of the lungfish have a sequence that it is also found in mammals but has less similarity to those of reptiles, amphibians, and ray-finned fish. In some species, this motif is also associated with repeats.

set with all protein coding genes combined was subjected to neighbor joining (NJ) (distance matrices were calculated based on Kimura distances), maximum parsimony (MP), and maximum likelihood (ML) analyses. This data set was also analyzed separately with MP, NJ, and ML by excluding third codon positions entirely and excluding transitions in third codon positions. (2) A set comprising all tRNA genes was subjected to MP, NJ,

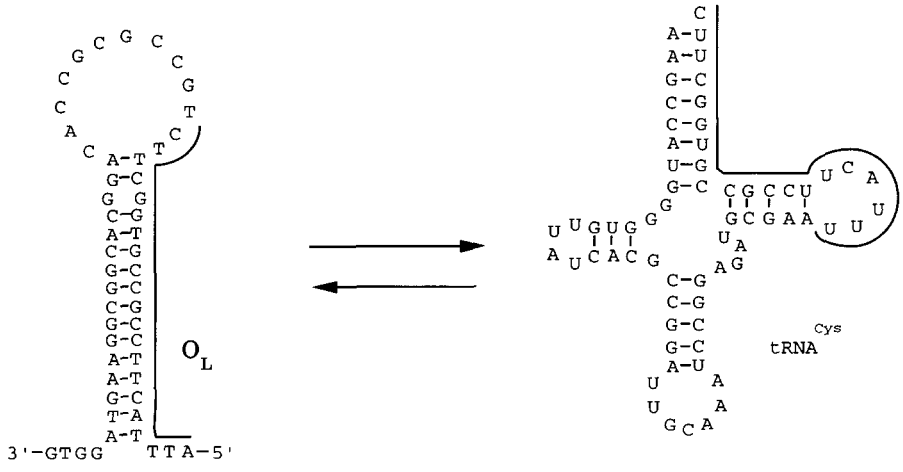

FIGURE 3.—Proposed stem-loop and cloverleaf secondary structures for the L-strand origin of replication and the tRNA^{Cys}, respectively. Both structures partially share the same sequence, which is underlined in both configurations.

TABLE 4

Overall base composition of the 13 protein-coding genes of fish, an amphibian and mammals

	Codon position	A	G	C	T
Lungfish	1	27.6	23.6	25.4	23.4
	2	18.4	13.3	27.2	41.1
	3	34.7	8.4	28.5	28.4
Frog	1	29.9	21.0	23.3	25.8
	2	20.5	11.6	27.2	40.7
	3	41.2	6.5	22.3	30.0
Trout	1	25.4	26.4	26.8	21.4
	2	18.2	13.8	27.7	40.3
	3	33.4	8.9	33.9	23.8
Carp	1	27.1	25.9	26.4	20.6
	2	18.5	14.0	28.2	39.3
	3	44.2	5.9	31.3	18.6
Loach	1	27.2	26.4	25.6	20.8
	2	18.5	13.7	27.7	40.1
	3	35.8	9.6	34.6	20.0
Lamprey	1	30.4	22.6	22.9	24.1
	2	19.0	12.9	26.5	41.6
	3	41.3	3.8	21.5	33.4
Mammals[a]	1	32.1	20.7	24.4	22.8
	2	19.5	12.2	26.2	42.1
	3	42.4	5.0	31.2	21.4

Values are percentages.
[a] JANKE *et al.* (1994).

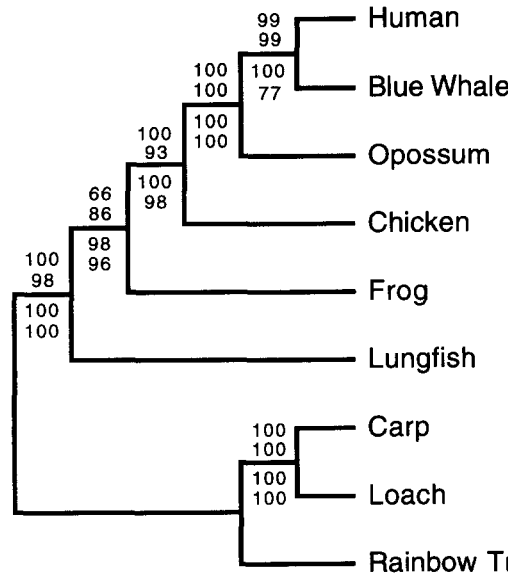

FIGURE 4.—Majority rule bootstrap (FELSENSTEIN 1985) consensus tree of vertebrates based on 100 replications. Two data sets were subjected to MP (bootstrap values above branches) and NJ (bootstrap values below branches) analyses. The first data set includes all mitochondrial protein coding genes combined (bootstrap values upper of each pair of numbers). The second data set comprises a combination of all mitochondrial tRNA genes (bootstrap values lower of each pair of numbers).

and ML phylogenetic analyses. In the tRNA data set all position, irrespective of secondary structure, were weighted equally. (3) Each protein coding gene was analyzed separately with MP, NJ, and ML. Analyses with all phylogenetic methods were also performed excluding third codon positions in each gene and in MP also third codon position transitions were excluded in separate analyses. Confidence levels for all neighbor joining and maximum parsimony analyses from all data sets were assessed by bootstrap analyses based on 100 replications (FELSENSTEIN 1985). In all MP analyses with PAUP (version 3.1.1), the heuristic search option was used.

Performance of lamprey as outgroup: Initially, all data sets were rooted using the lamprey mitochondrial DNA sequence (LEE and KOCHER 1995) as outgroup. Surprisingly, trees with odd topologies and low bootstrap values were obtained regardless of the phylogenetic method. This seemed especially surprising for the case of tRNAs, which have been used to infer phylogenetic relationships among vertebrates even using sea urchin as outgroup (KUMAZAWA and NISHIDA 1993), which diverged >600 mya (SIMMS *et al.* 1993). However, the analyses of KUMAZAWA and NISHIDA (1993) did not include any fish and when fish tRNA sequences were added to this data set unorthodox groupings resulted. Since lampreys diverged from the main vertebrate lineage ~550 mya (CARROLL 1988; LEE and KOCHER 1995), it is likely that multiple substitutions might have accumulated along the sequence, hindering the recovery of correct phylogenetic relationships among vertebrate species and impeding tree reconstruction. When lamprey was used as outgroup and fish were excluded from our tRNA data set, we were able to recover the well-established topology for relationships among vertebrates (data not shown).

Phylogenies based on combined tRNA and protein coding gene data sets: According to paleontological evidence, the separation of ray-finned fish from the lineage leading to lobe-finned fish occurred ~410 mya (CARROLL 1988). When trout, carp and loach were used as outgroups, and the combined sets of protein coding or tRNA genes were analyzed, all three phylogenetic methods yielded identical, congruent, and strongly supported topologies with the expected branching order (Figure 4). In these trees, the lungfish is unequivocally placed as the sister group of tetrapods.

Identical topologies were also obtained with all phylogenetic methods (MP, NJ, ML), when transitions in third codon positions of the combined protein coding genes were excluded from the analyses. However, when third codon positions were excluded completely from the analysis of this data set, NJ and ML methods yielded the expected topology whereas MP failed to recover it.

The robustness of these results was confirmed by the high bootstrap values obtained in NJ and MP trees (Figure 4). Interestingly, the bootstrap values yielded by the protein coding gene are higher than those obtained

TABLE 5

Confidence in maximum parsimony estimates

	All				No Ts in third				No third			
	Shortest	Expected	Δ	Percentage	Shortest	Expected	Δ	Percentage	Shortest	Expected	Δ	Percentage
ND1	1509	1515	6	0.4 ± 6.6	1115	1125	10	0.8	549	553	4	1.6 ± 3.1
ND2	2144	2150	6	0.3 ± 10.4	1749	1752	3	0.2	1031	1034	3	0.4 ± 7.4
COI	1937	1947	10	0.5 ± 11.4	1132	1135	3	0.3	385	388	3	0.8 ± 5.0
COII	985	999	14	1.4 ± 9.1	625	639	14	2.2	315	325	10	3.1 ± 5.1
ATPase 8	357	365	8	2.2 ± 5.2	307	305	2	0.6	206	211	5	2.3 ± 4.1
ATPase 6	1185	1204	19	1.6 ± 8.1	874	887	13	1.5	458	468	10	2.1 ± 4.5
COIII	1019	1031	12	1.2 ± 8.4	651	655	4	0.6	267	268	1	0.4 ± 2.2
ND3	624	643	19	3.0 ± 7.9	464	481	17	3.5	246	259	13	4.6 ± 4.8
ND4L	570	578	8	1.4 ± 6.0	449	462	13	2.8	264	273	9	3.3 ± 4.4
ND4	2515	2515	—	—	1867	1868	1	<0.1	1058	1061	3	0.3 ± 6.2
ND5	3372	3381	9	0.2 ± 12.9	2510	2520	10	0.4	1495	1495	—	—
ND6	1103	1116	13	1.2 ± 10.2	849	858	9	1.0	533	536	3	0.6 ± 4.3
Cyt*b*	1654	1665	11	0.6 ± 10.0	1135	1136	1	0.1	542	546	4	0.7 ± 6.0

The differences in length of the expected tree (Figure 4) and the shortest trees yielded by each gene (Δ) are indicated (in number of steps) with their standard deviation estimated by TEMPLETON's (1983) formula. Differences are also shown as percentage.

from the tRNA data set, contradicting the idea that only tRNA but not protein coding genes are able to estimate phylogenetic relationships among taxa that originated >300 mya (KUMAZAWA and NISHIDA 1993). A recent study on the origin of tetrapods (HEDGES *et al.* 1993), using complete mitochondrial rRNAs as data set and a ray-finned fish as outgroup seemed to support the position of lungfish as sistergroup to land vertebrates. Our results show that ray-finned fish are a reliable outgroup to assess relationships among tetrapods, leading to appropriate topologies when large data sets (combined tRNA, protein coding, or rRNA genes, see HEDGES *et al.* 1993) are assayed.

Phylogenetic performance of each mitochondrial protein coding gene: MP, NJ, and ML analyses such as those performed with the combined data sets were also carried out for each protein coding gene separately. This was done to elucidate which of the mitochondrial protein coding genes are able to recover the expected topology (Figure 4) and to determine which genes are appropriate for inferring deep branch phylogenies.

TABLE 6

Statistical confidence of maximum likelihood trees

	All		No 3rd	
	log *l*	Δl_i	log *l*	Δl_i
ND1	−7213.9	−9.9 ± 8.9	−3333.1	−7.8 ± 8.7
ND2	—	—	—	—
COI	−10120.5	8.8 ± 16.4	−3450.6	−5.4 ± 13.4
COII	−4788.1	−14.2 ± 9.1	−2105.8	−11.1 ± 10.2
ATPase 8	−1442.9	−2.1 ± 2.9	−886.4	3.7 ± 3.6
ATPase 6	−5440.3	−20.8 ± 10.8	−2633.0	−15.7 ± 8.5
COIII	−5252.8	−8.6 ± 18.4	−2047.0	−3.3 ± 4.9
ND3	−2659.8	−8.7 ± 5.0	−1350.9	−16.0 ± 8.6
ND4L	−2509.8	−14.8 ± 6.6	−1324.5	−9.4 ± 5.5
ND4	—	—	—	—
ND5	−15010.4	−14.7 ± 16.7	—	—
ND6	−4098.1	−1.6 ± 4.0	−2410.7	−4.3 ± 9.0
Cyt*b*	−8166.0	−6.2 ± 14.8	−3620.4	−7.4 ± 13.1

The differences in log-likelihood (Δl_i) between the best tree obtained for each gene and the expected tree (Figure 4) are shown with their standard error estimated by KISHINO and HASEWAGA'S formula (1989). The log-likelihood (log*l*) of the best tree for each gene is also indicated. Genes in bold are those for which the SE is larger than Δl_i and therefore the best tree is not significantly more likely than the expected tree (Figure 4). The ML trees based on ND2 and ND4 genes are the expected ones. When third positions are not included in the ML analyses, ND5 yields the expected tree as the best ML tree and, the best ML trees for ND1 and COII are not better than the expected topology (Figure 4).

TABLE 7

Phylogenetic relationships among vertebrates

Trees	Protein genes			tRNA genes
	All	No Ts in 3rd	No 3rd	All
1. (trout,(loach,carp),(lungfish,(frog,(chicken,(marsupial,(whale,human))))));	0.0[a] ± —	0.0[b] ± —	0.0[c] ± —	0.0[d] ± —
2. (trout,(loach,carp),((lungfish,frog),(chicken,(marsupial,(whale,human)))));	−59.9 ± 27.8	−109.8 ± 36.3	−51.4 ± 19.7	−18.8 ± 8.4
3. (trout,(loach,carp),((chicken,frog),(lungfish,(marsupial,(whale,human)))));	−176.1 ± 41.7	−172.1 − 53.3	−127.3 ± 36.5	−41.2 ± 14.7
4. (trout,(loach,carp),(frog,((lungfish,chicken),(marsupial,(whale,human)))));	−185.0 ± 40.8	−211.9 ± 53.3	−124.5 ± 35.9	−49.8 ± 15.2
5. (trout,(loach,carp),((lungfish,chicken),(frog,(marsupial,(whale,human)))));	−196.7 ± 40.8	−225.8 ± 52.2	−146.2 ± 35.4	−45.9 ± 14.7
6. (trout,loach,(carp,(frog,(lungfish,(chicken,(marsupial,(whale,human)))))));	−199.2 ± 44.8	−308.2 ± 60.5	−143.8 ± 36.6	−73.5 ± 15.8
7. (trout,(loach,carp),(chicken,(frog,(lungfish,(marsupial,(whale,human))))));	−206.0 ± 44.2	−231.3 ± 56.7	−159.7 ± 37.3	−56.9 ± 14.4
8. (trout,carp,(loach,(chicken,(frog,(lungfish,(whale,(marsupial,human)))))));	−570.5 ± 62.2	−708.4 ± 83.0	−441.7 ± 54.3	−120.1 ± 19.6
9. (trout,((loach,carp),(lungfish,(frog,chicken))),(human,(marsupial,whale)));	−734.3 ± 64.9	−825.5 ± 79.3	−696.7 ± 57.1	−74.6 ± 19.9
10. (trout,chicken,(carp,(loach,(lungfish,(frog,(marsupial,(whale,human)))))));	−738.1 ± 70.7	−863.3 ± 86.9	−692.0 ± 60.7	−154.2 ± 22.6

Differences in log-likelihood (Δl_i) between tree-*i* and the maximum likelihood tree and their standard error calculated by KISHINO and HASEWAGA's formula (1989) are shown. The alternative trees analyzed (2–10) are those given by the maximum likelihood method as best trees by some of the protein genes alone (2, ND1 all positions; 3, ND5 all positions; 4, ATPase 6 no 3rd positions; 5, ATPase 6 all position; 6, ND1 no 3rd positions; 7, ND3 no 3rd positions; 8, ND6 no 3rd positions; 9, ATPase 8, all positions; 10, COIII all positions).

[a] *Log l* = −89655.1.
[b] *Log l* = −74925.4.
[c] *Log l* = −43071.0.
[d] *Log l* = −7995.1.

Interestingly, with the exception of ND4 (with MP, NJ, and ML) and ND2 (with NJ, and ML, but not MP), none of the mitochondrial protein coding genes recovered by itself the correct branching order (Figure 4). This was the case even when transitions in the third codon position were excluded from the analysis or no third codon positions were considered at all. Furthermore, bootstrap values of the resulting trees were very low (data not shown).

In the parsimony analyses, only a few more steps are needed to recover the expected topology (Figure 4) from each gene, suggesting that the shortest tree obtained in each case is poorly supported and not statistically significantly different from the expected trees (Table 5). This finding from the MP analysis was confirmed with ML when standard errors of the difference in log-likelihood between the ML tree given by each gene and that of the correct tree were calculated by the formula of KISHINO and HASEWAGA (1989). This allowed us to evaluate whether the best tree was statistically significantly different estimate from the true tree (Table 6). All genes except *ATPase 6, ND3* and *ND4L* (among the fastest evolving mitochondrial protein coding genes; see LYNCH and JARRELL 1993) exhibited log-likelihood ratios for the expected tree that were not significantly lower than those of the best trees obtained in each case. This suggests that the expected tree cannot be statistically ruled out for most individual genes with the exception of *ATPase 6, ND3* and *ND4L*.

The same analysis (KISHINO and HASEWAGA 1989) was performed to evaluate the statistical support of the best tree (Figure 4) recovered from combined data sets (Table 7). In this case, since the best tree recovered was also the expected tree, we used the best topologies supported by individual protein genes (Table 6) as alternative trees. All of the alternative trees could be rejected since the difference in log-likelihood estimated in all cases was significantly different (Table 7). Presumably, the phylogenetic signal that every gene carries in its sequence, when combined is additive and strong enough to compensate for homoplasy contained in individual genes (the homoplasy of individual genes would be expected to be random).

The failure of most single mitochondrial protein genes to resolve relationships among the major groups of vertebrates, together with the successful behavior of all protein or tRNA genes combined, suggests that the limit for the utility of mitochondrial sequences might have been reached at ~400 million years. Our results might suggest that the level of homoplasy introduced by the lamprey mitochondrial DNA sequences is too high to be counteracted by the compensating effect of combining all mitochondrial protein-coding genes. However, it is unclear whether the lamprey mitochondrial genome is particularly homoplasious and that individual genes are therefore performing relatively poorly at this level of divergence.

It is not clear whether this result only applies to this particular study or whether it is general. Several reasons, such as differences in base composition, taxon sampling, differences in rates of evolution, and pronounced differences in branch lengths, and also short internodes, could account for this finding. The lack of resolution, especially for ancient nodes, is probably also due to extensive homoplasy in the data and the fact that relevant nodes, *i.e.*, the lungfish and amphibians lineages, originated within a narrow window in time of probably 20–30 million years, ~360 mya (reviewed in MEYER 1995). These reasons might be sufficient to constrain the phylogenetic resolving power of the phyloge-

netic methods, especially of maximum parsimony, and to hinder the recovery of the expected tree when each gene is analyzed individually. Mitochondrial genomes contain other information such as gene order (*e.g.*, BOORE and BROWN 1994; BOORE *et al.* 1995) that might permit phylogenetic inferences among lineages that diverged before the Devionian split of lungfishes and tetrapods.

We thank MOHAMMED ASLANI and GAIL MANDELL for assistance in the synthesis of oligonucleotide primers and MICHAEL LYNCH, GUILLERMO ORTI, and two referees for helpful comments and discussion. This work was partly supported by grants from the National Science Foundation to A.M. R.Z. is a postdoctoral fellow sponsored by a grant of the Ministerio de Educacion y Ciencia of Spain. The complete mtDNA sequence of the African lungfish has been deposited at the EMBL/GenBank data libraries under the accession number L42813.

LITERATURE CITED

ADACHI, J., and M. HASEWAGA, 1992 *MOLPHY: Programs for Molecular Phylogenetics I-PROTML: Maximum Likelihood Inference of Protein Phylogeny* (Computer Science Monographs, Vol. 27). Institute of Statistical Mathematics, Tokyo.

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. DE BRUIJN, A. R. COULSON *et al.*, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–464.

ANDERSON, S., M. H. DE BRUIJN, A. R. COULSON, I. C. EPERON, F. SANGER *et al.*, 1982 Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian genome. J. Mol. Biol. **156:** 683–717.

ARNASON, U., and A. GULLBERG, 1993 Comparison between the complete mitochondrial sequences of the blue and the fin whale, two species that can hybridize in nature. J. Mol. Evol. **37:** 312–322.

ARNASON, U., A. GULLBERG and B. WIDEGREN, 1991 The complete nucleotide sequence of the mitochondrial DNA of the Fin whale, *Balaenoptera physalus*. J. Mol. Evol. **33:** 556–568.

ARNASON, U., A. and E. JOHNSSON, 1992 The complete mitochondrial DNA sequence of the Harbor seal, *Phoca vitulina*. J. Mol. Evol. **43:** 493–505.

ARNASON, U., A. GULLBERG, E. JOHNSSON and C. LEDJE, 1993 The nucleotide sequence of the mitochondrial DNA molecule of the Grey seal, *Halichoerus grypus*, and a comparison with mitochondrial sequences of other true seals. J. Mol. Evol. **37:** 323–330.

ATTARDI, G., A. CHOMYN, R. F. DOOLITTLE, P. MARIOTTINI and C. I. RAGAN, 1986 Seven unidentified reading frames of human mitochondrial DNA encode subunits of the respiratory chain NADH dehydrogenase. Cold Spring Harbor Symp. Quant. Biol. **51:** 103–114.

BENTON, M. J., 1990 Phylogeny of the major tetrapod groups: morphological data and divergence dates. J. Mol. Evol. **30:** 409–424.

BISCHOFF, T. L. W. v., 1840 *Lepidosiren paradoxa. Anatomisch Untersucht und Beschrieben*. Leipzig.

BOORE, J. L., and W. M. BROWN, 1994 Mitochondrial genomes and the phylogeny of mollusks. Nautilus **2** (Suppl.): 61–78.

BOORE, J. L., T. M. COLLINS, D. STANTON, L. L. DAEHLER and W. M. BROWN, 1995 Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature **376:** 163–167.

BROWN, W. M., M. J. GEORGE and A. C. WILSON, 1979 Rapid evolution of mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971.

CARROLL, R. L., 1988 *Vertebrate Paleontology and Evolution*. W. H. Freeman, New York.

CHANG, Y. S., F. L. HUANG and T. B. LO,1994 The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. J. Mol. Evol. **38:** 138–155.

CLOUTIER, R., 1991 Patterns, trends, and rates of evolution within the actinistia, pp. 23–58 in *The Biology of Latimeria chalumnae and Evolution of Coelacanths*, edited by J. A. MUSICK, M. N. BRUTON and E. K. BALON. Kluwer, Hingham, MA.

DESJARDINS, P., and R. MORAIS, 1990 Sequence and gene organization of the chicken mitochondrial genome. J. Mol. Biol. **212:** 599–634.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DILLON, M. C., and J. M. WRIGHT, 1993 Nucleotide sequence of the D-loop region of the sperm whale (*Physeter-Macrocephalus*) mitochondrial genome. Mol. Biol. Evol. **10:** 296–305.

DODA, C. T., C. T. WRIGHT and D. A. CLAYTON, 1981 Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78:** 6116–6120.

FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

FELSENSTEIN, J., 1989 PHYLIP—phylogeny inference package (version 3.4.). Cladistics **5:** 164–166.

HEDGES, S. B., C. A. HASS and L. R. MAXSON, 1993 Relations of fish and tetrapods. Nature **363:** 501–502.

HIGGINS, D. G., and P. M. SHARP, 1989 Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. **5:** 151–153.

JAMA, M., Y. P. ZHANG, R. A. AMAN and O. A. RYDER, 1993 Sequence of the mitochondrial control region, tRNA (thr), tRNA (Pro) and tRNA (Phe) genes from the black rhinoceros, *Diceros bicornis*. Nucleic Acids Res. **21:** 4392–4392.

JANKE, A., G. FELDMAIER-FUCHS, K. THOMAS, A. VON HAESELER and S. PAABO, 1994 The marsupial mitochondrial genome and the evolution of placental mammals. Genetics **137:** 243–256.

JOHANSEN, S., P. H. GUDDAL and T. JOHANSEN, 1990 Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. Nucleic Acids Res. **18:** 411–419.

KISHINO, H., and M. HASEWAGA, 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.

KUMAZAWA, Y., and M. NISHIDA, 1993 Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. **37:** 380–398.

LEE, W. J., and T. D. KOCHER, 1995 Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. Genetics **139:** 873–887.

LYNCH, M., and P. E. JARRELL, 1993 A method for calibrating molecular clocks and its application to animal mitochondrial DNA. Genetics **135:** 1197–1208.

MACKAY, S. L. D., P. D. OLIVO, P. J. LAIPIS and W. W. HAUSWIRTH, 1986 Template-directed arrest of mammalian mitochondrial DNA synthesis. Mol. Cell Biol. **6:** 1261–1267.

MEYER, A., 1993 Evolution of mitochondrial DNA in fishes, pp. 1–38 in *Biochemistry and Molecular Biology of Fishes*, edited by P. W. HOCHACHKA and T. P. MOMMSEN. Elsevier Science, New York.

MEYER, A., 1995 Molecular evidence on the origin of tetrapods and the relationships of the coelacanth. Trends Ecol. Evol. **10:** 111–116.

MEYER, A., and S. I. DOLVEN, 1992 Molecules, fossils and the origin of tetrapods. J. Mol. Evol. **35:** 102–113.

MEYER, A., and A. C. WILSON, 1990 Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J. Mol. Evol. **31:** 359–364.

NAYLOR, G. J., T. M. COLLINS and W. M. BROWN, 1995 Hydrophobicity and phylogeny. Nature **373:** 555–556.

PANCHEN, A. L., and T. R. SMITHSON, 1987 Character diagnosis, fossils and the origin of tetrapods. Biol. Rev. **62:** 341–438.

PATTERSON, C., 1980 Origin of tetrapods: historical introduction to the problem, pp. 159–175 in *The Terrestrial Enviroment and the Origin of Land Vertebrates*, edited by A. L. PANCHEN. Academic Press, New York.

ROE, B. A., M. DIN-POW, R. K. WILSON and J. F. WONG, 1985 The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. **260:** 9759–9774.

ROMER, A. S., 1966 *Vertebrate Paleontology*. University of Chicago Press, Chicago.

ROSEN, D. E., P. L. FOREY, B. G. GARDINER and C. PATTERSON, 1981 Lungfishes, tetrapods, paleontolgy, and plesiomorphy. Bull. Am. Nat. Mus. Nat. Hist. **167:** 159–276.

SACCONE, C., G. PESOLE and E. SBISA, 1991 The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol. **33:** 83–91.

SEUTIN, G., B. FRANZ LANG, D. P. MINDELL and R. MORAIS, 1994 Evolution of the WANCY region in amniote mitochondrial DNA. Mol. Biol. Evol. **11:** 329–340.

SIMMS, M. J., A. S. GALE, P. GILLILAND, E. P. F. ROSE and G. D. SEVASTOPULO, 1993 Echinodermata, in *The Fossil Record 2*, edited by M. J. BENTON. Chapman and Hall, London.

SOUTHERN, S. O., P. J. SOUTHERN and A. E. DIZON, 1988 Molecular characterization of a cloned dolphin mitochondrial genome. J. Mol. Evol. **28:** 32–42.

STEINBERG, S., and R. CEDERGREN, 1994 Structural compensation in atypical mitochondrial tRNAs. Struct. Biol. **1:** 507–510.

STEWART, D. T., and A. J. BAKER, 1994 Patterns of sequence variation in the mitochondrial D-loop region of shrews. Mol. Biol. Evol. **11:** 9–21.

SWOFFORD, D. L., 1993 *PAUP: Phylogenetic Analysis Using Parsimony.* Illinois Natural History Survey, Champaign, IL.

TEMPLETON, A. R., 1983 Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution **37:** 221–244.

TZENG, C. S., C. F. HUI, S. C. SHEN and P. C. HUANG, 1992 The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. Nucleic Acids Res. **20:** 4853–4858.

WALBERG, M. W., and D. A. CLAYTON, 1981 Sequence and properties of the human KB cell and mouse L cell D-Loop regions of mitochondrial DNA. Nucleic Acids Res. **9:** 5411–5421.

WILKINSON, G. S., and A. M. CHAPMAN, 1991 Length and sequence variation in evening bat D-loop mtDNA. Genetics **128:** 607–617.

WOLSTENHOLME, D. R., 1992 Animal mitochondrial DNA: structure and evolution Int. Rev. Cytol. **141:** 173–216.

WONG, T. W., and D. A. CLAYTON, 1985 In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. Cell **42:** 951–958.

YOKOBORI, A. I., M. HASEWAGA, T. UEDA, N. OKADA, K. NISHIKAWA *et al.*, 1994 Relationship among coelacanths, lungfishes, and tetrapods: a phylogenetic analysis based on mitochondrial cytochrome oxidase I gene sequences. J. Mol. Evol. **38:** 602–609.

ZARDOYA, R., A. GARRIDO-PERTIERRA and J. M. BAUTISTA, 1995a The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss.* J. Mol. Evol. **41:** 942–951.

ZARDOYA, R., M. VILLALTA, M. J. LOPEZ-PEREZ, A. GARRIDO-PERTIERRA, J. MONTOYA *et al.*, 1995b Nucleotide sequence of the sheep mitochondrial DNA D-loop region and its flanking tRNA genes. Curr. Genet. **28:** 94–96.

Communicating editor: M. LYNCH