The Complete DNA Sequence of the Mitochondrial Genome of a "Living Fossil," the Coelacanth (*Latimeria chalumnae*)

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

ABSTRACT

The complete nucleotide sequence of the 16,407-bp mitochondrial genome of the coelacanth (*Latimeria chalumnae*) was determined. The coelacanth mitochondrial genome order is identical to the consensus vertebrate gene order which is also found in all ray-finned fishes, the lungfish, and most tetrapods. Base composition and codon usage also conform to typical vertebrate patterns. The entire mitochondrial genome was PCR-amplified with 24 sets of primers that are expected to amplify homologous regions in other related vertebrate species. Analyses of the control region of the coelacanth mitochondrial genome revealed the existence of four 22-bp tandem repeats close to its 3' end. The phylogenetic analyses of a large data set combining genes coding for rRNAs, tRNAs, and proteins (16,140 characters) confirmed the phylogenetic position of the coelacanth as a lobe-finned fish; it is more closely related to tetrapods than to ray-finned fishes. However, different phylogenetic methods applied to this largest available molecular data set were unable to resolve unambiguously the relationship of the coelacanth to the two other groups of extant lobe-finned fishes, the lungfishes and the tetrapods. Maximum parsimony favored a lungfish/coelacanth or a lungfish/tetrapod sistergroup relationship depending on which transversion:transition weighting is assumed. Neighbor-joining and maximum likelihood supported a lungfish/tetrapod sistergroup relationship.

COELACANTHS were for the first time recognized as a distinct taxonomic group >150 years ago by Louis Agassiz in his renowned book "*Recherches sur les Poissons Fossiles*" (1844). Since then, many more extinct coelacanth species have been discovered (CLOUTIER and FOREY 1991). These fossils range in age from the Early Devonian to the Late Cretaceous (HUXLEY 1861; WOODWARD 1891; STENSIÖ 1921). No fossil coelacanths are known from more recent times, and, hence, this group was believed to have gone extinct ~70–80 mya. However, in 1938, a living coelacanth (*Latimeria chalumnae*) was trawled off the estuary of the Chalumna river, near East London, South Africa (SMITH 1939, 1956; COURTENAY-LATIMER 1979). Since the sensational discovery of this living fossil, a total of only ~200 specimens have been caught off the Comoro archipelago near the eastern coast of Africa in the Indian ocean. It proved to be the only known coelacanth population (SCHLIEWEN *et al.* 1993; FRICKE *et al.* 1995).

Coelacanths are large gray-blue marine fishes whose morphology has remained seemingly unchanged during the last 400 million years (but see FOREY 1991). During the last 50 years, detailed anatomic and functional morphological analysis of the circulatory, reproductive, locomotive, and respiratory systems of the coelacanth have been conducted (*e.g.*, MILLOT 1954, 1955;

SMITH 1956; THOMSON 1966; BEMIS and NORTHCUTT 1991; MANGUM 1991; SCHULTZE and CLOUTIER 1991; TAMAI *et al.* 1994). These fishes are characterized by the possession of several distinct morphological, physiological, and life history traits, such as lobed paired fins, an heterocercal tail with an apical lobe, an intracranial joint, a rostral electroceptor organ, a urea retention mechanism and very large eggs (and give birth to living young) (*e.g.*, FOREY 1988; LONG 1995; MAISEY 1996). Unlike other sarcopterygians, the group to which they belong (see below), coelacanths lack choanae (internal nostril), maxillae and functional lungs.

Unfortunately, little is known about the coelacanth at a molecular level. To date, only the triosephosphate isomerase (KOLB *et al.* 1974), the α and β parvalbumin (PECHERE *et al.* 1978), and the α and β hemoglobin amino acid sequences (GORR *et al.* 1991), as well as the 18S rRNA (STOCK *et al.* 1991), 28S rRNA (HILLIS *et al.* 1991; ZARDOYA and MEYER 1996b), cytochrome *b* (MEYER and WILSON 1990), 12S rRNA (MEYER and WILSON 1990; MEYER and DOLVEN 1992; HEDGES *et al.* 1993), 16S rRNA (HEDGES *et al.* 1993), cytochrome oxidase subunit 1 (YOKOBORI *et al.* 1994), and Mhc class I (BETZ *et al.* 1994) nucleotide sequences have been reported from the coelacanth (reviewed by MEYER 1995).

Despite much prior effort, the phylogenetic position of the coelacanth still remains controversial (reviewed by MEYER 1995). It is well established that the coelacanth, together with lungfishes, and the extinct rhipidistians form the subclass Sarcopterygii, the lobe-

Corresponding author: Rafael Zardoya, Department of Ecology and Evolution, 662 Life Sciences, State University of New York, Stony Brook, New York 11794-5245. E-mail: rzardoya@life.bio.sunysb.edu

TABLE 1

PCR and sequencing primers used in the analysis of the coelacanth mitochondrial genome

PCR primers	Sequence (5′ → 3′)	Approximate product length (bp)
1. 16S F	AGT TCA GAC CGG AGC AAT CCA GG	600
2. LATI ND1 R	ACT TCG TAG GAA ATA GTT TGT GC	
3. LATI ND1 F1	CCT TGA TCG GGG CCC TTC GAG CA	600
4. LATI MET R	TCG GGG TAT GGG CCC GAA AGC TT	
5. LATI GLN F	TGA ACC TAT ACT AAA GAG ATC AA	750
6. LATI ND2 R	ATT TTT CGT AGC TGC GTT TGA TT	
7. LATI ND2 F	GGA CTT ATC CTG TCG ACT TGA CA	700
8. LATI TRP R	TTA AAG CTT TGA AGG CTC TTA GT	
9. LATI TRP F	AAG CAT CAT AAC CCT CAT AGC ACT	700
10. LATI COI R	GGC ATC ACT ATA AAG AAG ATT AT	
11. LATI COI F1	GAT GAC CAA ATT TAT AAT GTA GT	700
12. LATI COI R1	ATT GCC ATT ATC GCT CAG ACT AT	
13. LATI COI F2	AAG AAA GAA CCA TTC GGG TAT AT	1200
14. H7886[a]	TAN SWY CAR TAY CAY TGR TGN CC	
15. LATI COII F	TAA TTG ATG AAG TCG AAA ACC CTC A	650
16. LATI AT8 R	TTA GGC TCA TGG TCA GGT TCA	
17. LATI AT8 F	AGT GAA TGC CTC ARY TAA AYC C	1000
18. LATI COIII R	TTT TGT ACA GGT AGT GTG TGG TG	
19. LATI COIII F	ATA TAT CAA TGA TGA CGA GA	600
20. LATI COIII R1	ACA TCA ACG AAA TGT CAG TAC CA	
21. LATI COIII F1	TAC CAC TTC ACA TCA AAC CAC CA	1600
22. LATI ND4 R	TAG TAG GAC GGC GGC TAG TAC TAT	
23. LATI ND4 F1	CCT AAR GCC CAY GTA GAR GC	900
24. LATI LEU R	TTT GCA CCA AGA GTT TTT GGT CCT TA	
25. LATI LEU F	CTA AAG GAT AAT AGC TCA TCC ATT	600
26. LATI ND5 R	CAY CAG CCR ATT ART ARR AAT GAY AT	
27. LATI ND5 F	CAR YTA TTY ATC GGN TGR GAR GG	650
28. LATI ND5 R1	CCY ATY TTT CKG ATR TCY TGY TC	
29. LATI ND5 F2	TAA AGC AAT GCT GTT CCT ATG CT	850
30. LATI ND5 R3	AAN AGN GTN AGR TAN GTY TTR AT	
31. LATI ND5 F3	ATC AAC AAC TCG CAA CAA GGA C	1100
32. H15149[b]	AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CTC A	
33. LATI CYTB F	TAC TTA CAA AAA GAA ACC TG	850
34. LATI THR R	CGG CTT ACA AGG GCC GAT GCT TT	
35. LATI THR-F	AGC CTT AGT AGC TTA AAC CC	1000
36. LATI12S-R	AGC AAG GCT GGG ACC AAA CCT TT	
37. L-Phe	TAA AGC ATA GCA CTG AAA ATG	850
38. 12Sb	GAG GGT GAC GGG CGG TGT GT	
39. 12Sa	TGG GAT TAG ATA CCC CAC TAT	800
40. 16S-1H	KTA GCT CRC CYA GTT TCG GG	
41. 16S-1L	AGT ACC GCA AGG GAA ARC TGA AA	800
42. 16S-2H	GAT TRY GCT ACC TTY GCA CGG TCA	
43. 16Sar[c]	CGC CTG TTT ATC AAA AAC AT	500
44. 16Sbr[c]	CCG GTC TGA ACT CAG ATC ACG T	

TABLE 1

Continued

PCR primers	Sequence (5' → 3')	Approximate product length (bp)
45. LATI ND6-F	ATC CCC CCY AAA TAR ATC AAA AA	650
32. H15149[b]	AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CTC A	
46. LATI 16S-F	CAA TCC TTT CCA AGA GTC CAA AT	450
47. LATI 16S-R	GTA CTG GCT CCT TAA TGA ATA GT	
Sequencing		
LATI COI F3	CCC TGC TAT GAG CAC TAG GAT TT	
LATI ND3 F	AAT GTG GCT TTG ATC CTC TAG GA	
LATI ND3 R	TTT AGT CGC TCT GTT TGA TTA CC	

IUB code: R, A/G; Y, C/T; W, A/T; S, G/C; and K, G/T.
[a] YOKOBORI et al. (1994).
[b] KOCHER et al. (1989).
[c] PALUMBI et al. (1991).

finned fishes (ROMER 1966; CARROLL 1988; AHLBERG 1991; CLOUTIER and AHLBERG 1996) from which tetrapods were originated. The origin of land vertebrates from water-bound lobe-finned fishes involved both morphological and physiological innovations and was one of the most significant events in the history of vertebrates (PANCHEN and SMITHSON 1987). There is recent strong evidence supporting elpistostegids (or panderichthyids), a group of rhipidistians, as the extinct sister group of the earliest tetrapods (VOROBYEVA and SCHULTZE 1991; CLACK 1994; SCHULTZE 1994; AHLBERG et al. 1996; CLOUTIER and AHLBERG 1996).

However, there is no general agreement about the relationships among the coelacanth, lungfishes and the rhipidistian+tetrapod clade (e.g., SCHULTZE 1994; for review, see MEYER 1995). Since molecular phylogenetic data can only be collected from extant species, they can contribute only in limited ways to this debate (MEYER and WILSON 1990; MEYER and DOLVEN 1992; HEDGES et al. 1993; YOKOBORI et al. 1994; ZARDOYA and MEYER 1996b; see MEYER 1995 for a review). However, the establishment of the phylogenetic relationships among

the three extant sarcopterygian groups, i.e., lungfishes, the coelacanth and tetrapods, can be accomplished by the evolutionary analyses of molecular data sets and would help to recreate the sequence of evolutionary events that might have preadapted for the extinct sarcopterygian conquest of land. All three possible competing hypotheses that explain the relationships among living representatives of sarcopterygians, have been supported at various times by either morphological and/ or molecular data: (1) the lungfishes as the closest living relatives of tetrapods (ROSEN et al. 1981; FOREY 1987, 1988; MEYER and WILSON 1990). (2) The coelacanth as the sister group of tetrapods among living vertebrates (FRITZSCH 1987; SCHULTZE 1987; GORR et al. 1991). And (3) lungfishes and the coelacanth form a monophyletic group and, hence, are equally closely related to tetrapods (NORTHCUTT 1987; FOREY 1988; CHANG 1991; SCHULTZE 1994; ZARDOYA and MEYER 1996b). The apparently rapid successive origin and subsequent diversification of the three remaining lineages of sarcopterygians back in the Devonian makes the determination of their phylogenetic relationships difficult. Many mor-

Latimeria chalumnae mtDNA

16,407 bp

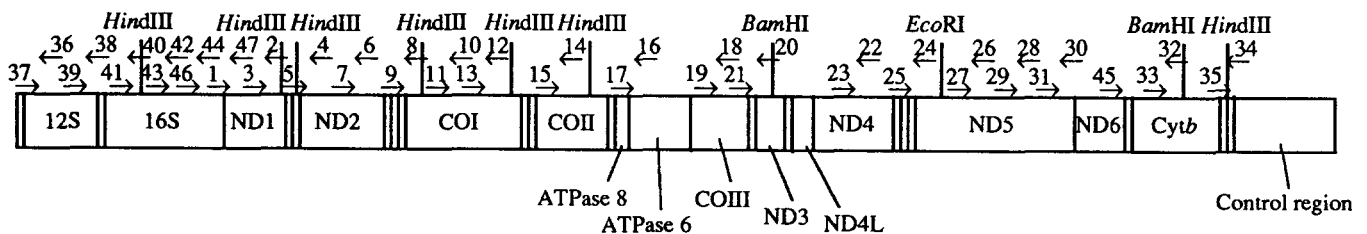

FIGURE 1.—Gene organization, restriction map, and cloning/sequencing strategy for the coelacanth mitochondrial genome. All protein coding genes are encoded by the H strand with the exception of ND6, which is coded by the L strand. Localization and direction of the primers used in the PCR amplification are denoted by arrows (see Table 1 for the primer DNA sequence associated with each number).

tRNA-Phe→

12s rRNA→

```
   1 GTGAGTGTAGCTTAATAAAAAGCATAGCACTGAAGATGCTAAGATGAATTACAAAATATTCCACTGACACAAAGGTTTGGTCCCAGCCTTGCTATCAATT  100

 101 TTAACCAGGATTACACATGCAAGCATCAACTCCCCAGTGAGAATGCCCCTGACTTATCCGTCAAAGATAACAGGGAGTAGGTATCAGGCACACAACATTA  200

 201 ACGCTAGCCCAAGACACCTTGTCCAGCCACACCCCCAAGGGAACTCAGCAGTGATAGACATTGAATAATAAGTGAAAACTTGACTCAGCCATGGTTACAA  300

 301 GGGCCGGTCAACTCCGTGCCAGCCACCGCGGTTACACGGAAGACCCAAAATGATAACACTACCGGCGTAAAGCGTGATTAAAGGACACCCACCATAATGG  400

 401 AGCCACAAATAACTAAAGCTGTTATACGCACTTAAAAAAATATGCTCATCACACGAAAGTAACTCCAGCACCCAAAGGAACCCTGAACCCACGAAAGCTA  500

 501 AGAAACAAACTGGGATTAGATACCCCACTATGCTCAGCCCTAAACACAAACAATTCAAACACACACTGTTCGCCAGGGGAACTACAAGCGCCAGCTTCAA  600

 601 ACCCAAAGGACTTGGCGGCACTTCAAACCCACCTAGAGGAGCCTGTTCTAAAACTGACAACCCCCACCTAACCTCACCATCCCTAGCCATTAAACCAGCC  700

 701 TATATACCGCCGTCGCCAGCCCACCCTGTGAAGGAAATACAATGGGCAAAATAAAAAAATTAAAAACGTCAGGTCGAGGTGTAGCAAATGAGATGGGAA  800

 801 GAAATGGGCTACATTTTCTAAATATAGAATATTACGAAAAAATACAGCGAAACCTGTACTTTGAAGGAGGATTTAGCAGTAAAAGGGGAATAGAGAGCCC  900

 901 CTCTGAAACCGGCCCTGAAATGCGCACACACCGCCCGTCACTCTCCTCACCCAAAATCGGCCCCATCTTTTAATAAACAAAAAACCAAGCATACAAGTAG 1000
```

tRNA-Val→

```
1001 AGGAGGCAAGTCGTAACAAGGTAAGTGTACCGGAAGGTGCACTTGGACTAATCAAAATGTAGTTTAACAAAACACCTCCCTTACACTGAGAAGACACCCA 1100
```

16S rRNA→

```
1101 CGAAAGCGGGTCATTTTGAGCTATATAGCTAGGCCAACAAAACCATAACCACTACACCATTAACCAAAAACTTGAACAAACAAAACCTAAACCATTTACC 1200

1201 ATCCAAGTATAGGCGATAGAAAAGACACCAGGCACAATAGTAAAAGTACCGCAAGGGAACGCTGAAAAGAAATGAAACAACTCGTAAAAGCAATAAAAA 1300

1301 CCAAAGACTAACCCTTGTACCTTTTGCATTATGATCTAGTTAGACCCCACCGGGCAAAATGAATTTAAGTCCAACCCCCCGAAACTAAGTGAGCTACTTC 1400

1401 GAAACAGCCTATGAGGGCAAACCCTTCTCTGTGGCAAAAGAGTGGGAAGATTTCCAAGTAGAGGCGATAAACCTAACGAGCTTAGTGATAGCTGGTTATT 1500

1501 CGAGAAATGATTTTTAGTTCAGCCCCTCGGCCTCCAAACCCTAACAACATAAAACCAATGTGAGCCAGAGAGAGTTATTCAAAGGGGGTACAGCCCCTTT 1600

1601 GAAAAAGGACACAACCTTCTAAGACAGGATAAAGACTATATTAACTCAAGGATCAGCTGCCCAAGTGGACCTAAAAGCAGTCACCTAAAAAAGCGTTAAA 1700

1701 GCTTAAGCAGCACTAAACCAACCTATACTAATAAATTTAACCTCATTCCACCATCACTATCGAATTATTCTATATGCATAGAAGAATAAATGCTAGAATT 1800

1801 AGTAACAAGAAGGCCATTAAGCCTTCTCTAACTGCATAAGTGTACATCAGATTAGATTAACCACTGATAATTAACGACTTCAAAGAGAATACTATGACAT 1900

1901 AAAACAAGAAAAGCACACACGCCCACATCGTTAATCCAACACAGGAATGCAACCACGAAAGATTAAAAGAAAGAAAAGGAACTCGGCAAACTATAAGCCC 2000

2001 CGCCTGTTTACCAAAAACATCGCCTCCCGCCAACAACAGAAGTATTGGAGGTCCCGCCTGCCCAGTGACAAGATTTAACGGCCGCGGTATCCTGACCGTG 2100

2101 CAAAGGTAGCGAAATCACTTGTCTTTTAAATGAAGACCTGTATGAATGGCACCACGAGGGCTTAACTGTCTCCTCTTTCCAATCAGTAAAATTGATCTGT 2200

2201 CCGTGCAGAAGCGGACTTAACTACATTAGACGAGAAGACCCTGTGGAGCTTCAGACACAAAGCCAACTACAAATAAACACTTAGATAATAAATTGATAAA 2300

2301 ACCAGTAGCCTAATACTGGCCCTATTGTCTTTGGTTGGGGCGACCACGGAGAAAAAACAATCCTCCAAGCCGATTGGTACCACCTGTACTAAAACAAAGG 2400

2401 GTAACACCCCAAAGTAATAAAAATTTTATCGGACATGACCCAGGGACTAAACCTGATCAACGAACCAAGTTACCCCAGGGATAACAGCGCAATCCTTTCC 2500

2501 AAGAGTCCAAATCGACGAAAGGGTTTACGACCTCGATGTTGGATCAGGACACCCCAATGGTGAAGCCGCTATTAAAGGTTTGTTTGTTCAACAATTAAAG 2600

2601 TCCTACGTGATCTGAGTTCAGACCGGAGAAATCCAGGTCAGTTTCTATCTATGATGTTAATTCTCCCAGTACGAAAGGACCGGAGAACTTGAGCCAATGC 2700
```

tRNA-Leu(UUR)→

```
2701 CACAAGCACGCTCATCTTCAACCTGATGAAAACAACTAAAACAGATAAAGAAGAACACAACCGCTCCCGTAAACAACGGGCAATGCCGGGGTGGCAGAGC 2800
```

NADH 1→

```
                                 M  T  K  I  I  T  H  L  L  N  P  L  A  V
2801 CTGGCTAAATGCAAGAGGCCTAAACTCTCTCTCTCAGAAGTTCAAATCCTCTCCCCAGCTATGACAAAAATTATTACACACCTGCTAAACCCACTAGCTG 2900
        I  I  P  I  L  L  A  V  A  F  L  T  L  I  E  R  K  V  L  G  Y  M  Q  L  R  K  G  P  N  I  V  G  P
2901 TCATCATTCCAATTTTACTAGCCGTAGCGTTCCTGACACTCATCGAACGAAAAGTACTAGGTTATATGCAACTACGAAAGGGTCCTAATATCGTAGGCCC 3000
        Y  G  L  L  Q  P  L  A  D  G  L  K  L  F  I  K  E  P  V  R  P  S  T  S  S  P  L  L  F  I  T  T  P
3001 ATATGGTCTCCTACAACCCCTAGCAGATGGACTAAAACTATTCATTAAGGAGCCAGTACGACCTTCTACATCATCCCCGCTACTCTTCATTACCACTCCA 3100
        M  L  A  L  T  M  A  L  T  W  L  P  L  P  L  P  H  P  M  T  N  L  N  L  G  M  L  F  I  L  A  I  S
3101 ATACTAGCACTCACCATAGCACTAACTCTATGACTACCACTACCCCTCCCACATCCAATAACAAACCTAAACCTAGGAATATTATTTATCCTAGCAATCT 3200
        S  L  T  V  Y  S  I  L  G  S  G  W  A  S  N  L  K  Y  A  L  I  G  A  L  R  A  V  A  Q  T  I  S  Y
3201 CAAGTCTAACAGTATACTCAATTCTAGGCTCCGGCTGAGCATCAAACTTAAAATATGCCCTAATTGGGGCCCTCCGAGCAGTCGCACAAACAATCTCCTA 3300
        E  V  S  L  G  L  I  L  L  A  M  I  I  F  A  G  G  F  T  L  T  T  F  N  T  S  Q  E  T  I  W  L  L
3301 TGAAGTAAGCCTAGGACTTATCCTACTGGCCATAATCATCTTCGCAGGCGGTTTTACACTAACAACATTTAATACATCACAGGAAACCATTTGACTTCTA 3400
        T  P  G  W  P  L  A  A  M  W  Y  I  S  T  L  A  E  T  N  R  A  P  F  D  L  T  E  G  E  S  E  L  V  S
3401 ACGCCAGGATGACCACTCGCAGCAATATGATACATCTCAACCTTAGCAGAAACCAACCGAGCCCCATTTGACCTCACAGAAGGAGAATCAGAACTTGTAT 3500
        G  F  N  V  E  Y  A  G  G  P  F  A  L  F  F  L  A  E  Y  A  N  I  L  L  M  N  T  L  S  T  I  L  F
3501 CGGGGTTTAATGTGGAATATGCAGGAGGACCATTTGCACTATTCTTTCTAGCAGAATATGCAAATATTCTACTAATAAACACACTATCAACTATTCTATT 3600
        M  G  A  M  H  N  P  I  T  P  E  L  T  S  I  N  L  M  I  K  A  S  A  L  S  M  L  F  L  W  V  R  A
3601 TATAGGAGCCATACACAACCCAATCACACCAGAACTAACCTCAATTAACCTAATGATTAAAGCCTCCGCACTATCAATACTCTTCCTATGAGTACGAGCC 3700
```

FIGURE 2.—Complete L-strand nucleotide sequence of the coelacanth mitochondrial genome. Position 1 corresponds to the first nucleotide of the *tRNA*^Phe gene. Direction of transcription for each gene is represented 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). Complete termination codons are indicated (*). tRNA genes are underlined and the corresponding anticodons are overlined. In the control region, four repeats are shown.

```
         S  Y  P  R  F  R  Y  D  Q  L  M  H  L  V  W  K  N  F  L  P  I  T  L  A  M  I  L  W  H  T  S  L  P  I
3701  TCATACCCACGATTCCGATATGACCAGCTAATACACTTAGTATGAAAAAACTTCTTACCAATCACCCTAGCCATGATCCTATGACACACCTCCCTTCCAA  3800
         F  T  G  S  L  P  P  Q  T  *   tRNA-Ile→
3801  TTTTTACAGGAAGCTTACCACCACAAACCTAAAGGAAACGTGCCCGAATAAACAAGGACCACTTTGATAGAGTGGACTATAGGAGTTAAAACCTCCTCGC  3900
                                                            ←tRNA-Gln

                                                            tRNA-Met→
3901  TTCCTAGGAAAATAGGACTTGAACCTATACTAAAGAGATCAAAACTCTTGGTGCTACCACTACACCATCTCCTAGTAAAGTCAGCTAAAAAAGCTTTCGG  4000
                                           NADH 2→
                                           M  S  P  Y  V  T  M  I  L  I  S  S  L  G  L  G  T  T  I
4001  GCCCATACCCCGAACATGTTGGTTAAACCTCCCTCCTTTACTAATGAGCCCTTACGTAACAATAATCCTTATCTCAAGCCTTGGACTCGGGACAACAATT  4100
         T  F  T  S  S  S  W  L  M  A  W  M  G  L  E  I  N  T  L  A  I  T  P  L  M  V  K  Q  H  H  P  R  A  T
4101  ACATTTACAAGCTCATCCTGACTGATAGCTTGAATAGGTCTAGAAATTAATACCCTAGCCATCACCCCCCTAATAGTAAAACAACATCACCCTCGGGCAA  4200
         E  A  T  T  K  Y  F  L  T  Q  A  T  A  S  G  L  L  L  F  A  T  L  N  N  A  W  M  T  G  E  W  N  T
4201  CTGAAGCCACAACAAAATATTTTCTTACCCAGGCAACAGCATCAGGACTGCTATTATTCGCAACCCTTAACAACGCTTGGATAACAGGAGAATGAAACAC  4300
         M  E  L  S  N  N  L  S  A  P  M  I  T  M  A  L  A  L  K  M  G  V  A  P  M  H  F  W  L  P  E  V  L
4301  AATAGAACTATCAAACAATTTATCCGCCCCAATAATTACAATAGCCCTCGCACTAAAGATAGGAGTAGCACCAATACACTTCTGATTACCAGAAGTGCTA  4400
         Q  G  L  P  L  L  T  G  L  I  L  S  T  W  Q  K  L  A  P  F  T  L  L  Y  M  T  S  H  E  L  N  T  T  T
4401  CAAGGACTCCCCCTACTAACTGGACTTATCCTGTCGACTTGACAAAAACTAGCCCCCTTCACCCTACTATATATAACATCGATGAATTAAACACAACAA  4500
         M  T  I  L  G  L  T  S  T  I  G  G  L  G  G  N  Q  T  Q  L  R  K  V  L  A  Y  S  S  I  A  H
4501  CAATAACAATCCTAGGATTGACATCAACAATTATCGGTGGCCTTGGTGGATTAAACCAAACTCAACTGCGAAAAGTCCTAGCTTACTCATCAATTGCACA  4600
         L  G  W  M  V  I  I  I  Q  Y  S  K  T  L  A  L  L  N  L  L  L  Y  I  T  M  T  S  T  A  F  L  T  L
4601  CCTCGGATGAATAGTTATCATTATCCAATACTCCAAAACACTAGCCCTACTAAACCTACTGCTATATATTACAATAACATCAACAGCATTTTTAACACTT  4700
         M  T  L  S  A  T  K  I  N  T  L  S  T  K  W  A  T  T  P  I  A  T  M  T  A  M  L  A  L  L  A  L  G  G
4701  ATGACTCTATCAGCCACAAAAATTAATACCCTATCAACAAAATGAGCAACAACCCCTATCGCAACTATAACTGCAATACTAGCTCTACTAGCATTAGGAG  4800
         L  P  P  L  T  G  F  M  P  K  W  L  I  L  Q  E  L  T  K  Q  N  L  P  A  L  A  T  L  M  A  L  S  A
4801  GTCTCCCACCACTAACAGGATTTATACCAAAATGACTAATTTTACAAGAACTTACCAAGCAAAATTTACCCGCTCTAGCCACACTAATAGCCCTATCAGC  4900
         L  L  S  L  F  F  Y  L  R  M  C  H  T  M  T  L  T  I  S  P  N  T  N  N  N  M  I  T  W  R  K  K  P
4901  CCTATTAAGCCTCTTCTTCTACCTACGAATATGCCATACAATAACCCTTACAATCTCACCAAACACAAACAATAACATAATTACATGACGAAAGAAACCT  5000
         G  Q  K  A  L  P  L  A  M  L  S  I  M  T  L  M  A  L  P  T  T  P  T  M  V  A  I  M  N  *  tRNA-Trp
5001  GGCCAGAAAGCACTACCCCTAGCCATACTAAGCATCATAACCCTCATAGCACTCCCAACAACCCCAACAATAGTAGCCATCATAAACTAATAAGAGACTT  5100
         →
5101  AGGAATAAACCAAGACCAAGAGCCTTCAAAGCCCTCAGCAGGAGTGAAAACCTCCTAGTCTCTGGATAAGACCTGTAGGATACTACCCCACATATTCTGA  5200
                              ←tRNA-Ala
5201  ATGCAAACCAGACACTTTAATTAAGCTAAGGCCTCCACTAGATGGGTAGGCCTCGATCCTACAAAATATTAGTTAACAGCTAAAAGCCCAAACCGGCGAG  5300
      ←tRNA-Asn                                                                                    ←tRNA-
5301  CATCCATCCTAACATTCCAAGAAAAAAAGGAATGTTTGAAGCCTCAACAAGCACCACTTGTCCCTCCAGGTTTGCAACCTGACATGCTAAACACCATGAG  5400
                                                                            COI→
      Cys                                                   ←tRNA-Tyr M  M  I  T  R  W  L  F
5401  GCTTGGTAAGAGGAGGAATTGAACCTCCCTACACGGGGCTACAACCCGCCGCTAAGAGCAATCAGCCATCTTACCTGTGATAATCACTCGTTGACTATTC  5500
         S  T  N  H  K  D  I  G  T  L  Y  M  I  F  G  A  W  A  G  M  V  G  T  A  L  S  L  L  I  R  A  E  L  S
5501  TCAACCAACCATAAAGACATTGGTACCCTATACATGATCTTCGGTGCCTGAGCTGGAATAGTTGGAACCGCCCTAAGCCTGCTTATTCGAGCTGAACTCA  5600
         Q  P  G  A  L  L  G  D  D  Q  I  Y  N  V  V  V  T  A  H  A  F  V  M  I  F  F  M  V  M  P  I  M  I
5601  GCCAACCTGGGGCTCTCCTGGGCGATGACCAAATTTATAATGTAGTCGTTACAGCACATGCATTCGTGATAATCTTCTTTATAGTAATACCGATCATAAT  5700
         G  G  F  G  N  W  L  I  P  L  M  I  G  A  P  D  M  A  F  P  R  M  N  N  M  S  F  W  L  L  P  P  S
5701  CGGCGGGTTTGGCAACTGATTAATTCCCCTGATGATTGGGGCACCCGACATAGCATTTCCACGTATAAACAACATAAGCTTCTGACTACTACCACCCTCA  5800
         L  L  L  L  L  A  S  S  G  V  E  A  G  A  G  T  G  W  T  V  Y  P  P  L  A  G  N  L  A  H  A  G  A  S
5801  CTCCTACTCCTACTAGCATCTTCTGGAGTAGAAGCAGGAGCAGGCACAGGATGGACAGTATACCCTCCACTAGCGGGCAACCTCGCCCATGCAGGAGCAT  5900
         V  D  L  T  I  F  S  L  H  L  A  G  V  S  S  I  L  G  A  I  N  F  I  T  T  V  I  N  M  K  P  P  T
5901  CCGTAGATTTAACAATTTTCTCCTTACATCTAGCCGGTGTATCCTCAATCTTAGGGGCCATCAACTTCATCACAACAGTAATCAACATAAAACCCCCAAC  6000
         M  T  Q  Y  Q  T  P  L  F  I  W  S  V  L  V  T  A  V  L  L  L  L  S  L  P  V  L  A  A  G  I  T  M
6001  AATAACACAGTATCAGACACCACTATTTATCTGATCAGTCTTAGTGACCGCCGTACTACTCCTACTCTCGCTACCGGTGCTAGCTGCCGGAATTACCATA  6100
         L  L  T  D  R  N  L  N  T  T  F  F  D  P  A  G  G  G  D  P  I  L  Y  Q  H  L  F  W  F  F  G  H  P  E
6101  CTACTGACAGATCGAAATCTAAACACAACATTCTTTGACCCTGCTGGAGGAGGAGACCCTATTCTATACCAACACCTATTCTGATTCTTCGGCCATCCTG  6200
         V  Y  I  L  I  L  P  G  F  G  M  I  S  H  I  V  A  Y  Y  S  G  K  K  E  P  F  G  Y  M  G  M  V  W
6201  AAGTATACATCCTAATTTTACCAGGATTTGGTATAATCTCACACATTGTGGCCTACTACTCTGGAAAGAAAGAACCATTCGGGTATATAGGTATAGTATG  6300
         A  M  M  A  I  G  L  L  G  F  I  V  W  A  H  H  M  F  T  V  G  M  D  V  D  T  R  A  Y  F  T  S  A
6301  AGCTATAATGGCAATTGGACTTCTAGGCTTCATCGTATGAGCCCATCATATATTTACCGTAGGAATGGATGTTGACACACGAGCATACTTTACATCAGCA  6400
         T  M  I  I  A  I  P  T  G  V  K  V  F  S  W  L  A  T  L  H  G  G  V  T  K  W  D  T  P  L  L  W  A  L
6401  ACCATAATTATTGCCATCCCAACAGGAGTAAAAGTGTTCAGCTGACTAGCGACACTTCACGGAGGAGTGACCAAATGAGACACACCCCTGCTATGAGCAC  6500
         G  I  F  L  F  T  V  G  G  L  T  G  I  V  L  A  N  S  S  L  D  I  I  L  H  D  T  Y  Y  V  V  A
6501  TAGGATTTATCTTTCTTTTTACAGTAGGAGGCCTAACAGGCATCGTACTGGCAAACTCATCACTAGACATCATCCTACATGACACTTATTACGTAGTAGC  6600
         H  F  H  Y  V  L  S  M  G  A  V  F  A  I  M  G  G  L  V  H  W  F  P  L  M  T  G  Y  T  L  H  N  T
6601  ACACTTCCACTATGTCCTATCAATAGGAGCAGTATTTGCAATCATAGGGGGACTCGTGCACTGATTTCCACTAATAACAGGATATACCTTACACAACACA  6700
```

FIGURE 2.—*Continued*

```
        W  T  K  I  H  F  G  V  M  F  T  G  V  N  L  T  F  F  P  Q  H  F  L  G  L  A  G  M  P  R  R  Y  S  D
6701  TGAACAAAAATCCACTTTGGTGTAATATTCACAGGAGTAAACCTAACATTTTTCCCACAACACTTCCTCGGACTAGCAGGAATACCACGACGTTACTCAG  6800
        Y  P  D  A  Y  T  L  W  N  T  V  S  S  I  G  S  L  I  S  L  I  A  V  I  M  F  M  F  I  L  W  E  A
6801  ACTATCCAGATGCCTATACTTTATGAAACACAGTATCATCAATTGGCTCTCTAATTTCACTAATTGCCGTAATCATATTTATATTTATCCTGTGAGAAGC  6900
        F  S  A  K  R  E  V  L  I  V  E  M  T  T  T  N  V  E  W  L  H  G  C  P  P  P  H  H  T  Y  E  E  P
6901  TTTCTCTGCCAAACGAGAAGTACTAATTGTAGAAATAACAACAACAAATGTAGAATGGCTGCACGGATGCCCACCACCACACCACACCATATGAAGAACCA  7000
        A  F  V  Q  A  P  R  *                                                              ←tRNA-Ser (UCN)
7001  GCATTCGTACAAGCTCCTCGATAAAACACTGAGAAAAGAGGGAATCGAACCCCCGTCAGCTAATTTCAAGACAGCCGCATCACCACTCTGCCATTTTCTT  7100
                                                                                                  COII→
      tRNA-Asp→                                                                            M  A  H  P  S  Q  L
7101  TATTAAGATTCTAGTAAAACAAATTACACTACCTTGTCAAGATAGAATTATGAGTGAAAACCTCATGAACCTTGACCAAATGGCACACCCATCACAGTTA  7200
        G  L  Q  D  A  A  S  P  V  M  E  E  L  L  H  F  H  D  H  A  L  M  I  V  F  L  I  S  T  L  V  F  Y  I
7201  GGATTACAAGATGCAGCTTCTCCCGTTATAGAAGAACTCCTCCACTTTCACGATCATGCACTAATAATTGTATTTTTAATTAGCACATTAGTATTTTACA  7300
        I  L  A  M  M  T  T  K  M  T  D  K  Y  I  L  D  A  Q  E  I  E  I  V  W  T  L  L  P  A  I  V  L  I
7301  TTATTCTAGCCATAATAACAACAAAAATAACTGACAAATATATCTTAGACGCACAAGAAATTGAAATTGTGTGAACACTACTCCCAGCAATCGTCCTAAT  7400
        L  V  A  L  P  S  L  R  I  L  Y  L  I  D  E  V  E  N  P  H  L  T  I  K  A  M  G  H  Q  W  Y  W  S
7401  CCTAGTTGCCCTACCCTCGCTACGAATCCTATATCTAATTGATGAAGTCGAAAACCCTCACCTAACAATTAAAGCAATAGGCCACCAATGATACTGAAGC  7500
        Y  E  Y  T  D  Y  E  E  L  S  F  D  S  Y  M  T  P  L  Q  D  L  N  P  G  Q  F  R  L  L  E  T  D  H  R
7501  TATGAGTACACGGACTATGAAGAACTAAGCTTCGACTCATACATAACACCACTACAAGACCTAAACCCGGGCCAATTCCGCTTGCTGGAAACAGACCATC  7600
        M  V  I  P  M  E  S  L  I  R  V  L  I  S  A  E  D  V  L  H  S  W  A  V  P  A  L  G  V  K  M  D  A
7601  GAATGGTTATCCCAATAGAGTCGCTTATCCGAGTACTAATTTCAGCTGAAGACGTACTACACTCATGAGCAGTCCCAGCCCTAGGAGTAAAAATAGATGC  7700
        V  P  G  R  L  N  Q  I  T  F  M  I  S  R  P  G  L  Y  Y  G  Q  C  S  E  I  C  G  A  N  H  S  F  M
7701  AGTCCCAGGGCGACTCAACCAAATTACATTCATAATTTCCCGACCAGGACTATATTATGGACAATGCTCAGAGATTTGTGGAGCAAACCACAGCTTTATA  7800
        P  I  V  L  E  A  I  P  L  D  P  F  E  D  W  S  S  M  L  E  E  A  tRNA-Lys→
7801  CCCATCGTACTTGAAGCAATCCCACTAGACCCCTTCGAAGACTGATCTTCATCAATGCTGGAAGAAGCCTCACTGAGAAGCTAAATAGAAAGCGATAGCC  7900
                                                      ATPase 8→
                                                      M  P  Q  L  N  P  S  P  W  L  L  I  L  L  F  S  W  L  I
7901  TTTTAAGCTAGAGACTGGTGAAAACAAACCCACCCTCAGTGCCATGCCACAACTAAACCCCTCCCCCTGACTACTAATCCTGCTATTCTCCTGACTCATC  8000
        F  L  T  M  L  P  S  K  T  Q  L  H  T  F  P  N  M  P  S  T  Q  N  M  C  K  Q  E  P  E  P  W  T  W  P
8001  TTCTTAACTATACTCCCCTCTAAGACACAATTACACACCTTCCCAAACATGCCATCAACACAAAATATATGCAAACAAGAACCAGAACCATGAACCTGAC  8100
      ATPase 6→
        M  S  L  N  F  F  D  Q  F  M  S  P  T  L  L  G  V  P  L  I  A  V  A  M  M  F  P  W  T  L  L  P  T
        W  A  *
8101  CATGAGCCTAAACTTCTTTGACCAATTTATGAGCCCAACACTATTAGGAGTACCACTCATTGCTGTAGCAATAATATTCCCATGGACCCTATTACCAACA  8200
        P  T  N  R  W  L  N  N  R  T  L  T  L  Q  N  W  F  I  G  R  F  T  N  Q  L  L  Q  P  L  N  T  G  G  H
8201  CCAACCAACCGATGACTTAATAACCGAACACTAACACTACAAAACTGATTTATCGGCCGCTTCACTAATCAACTACTACAACCATTAAACACTGGAGGAC  8300
        K  W  A  M  I  L  M  S  L  N  L  L  G  L  L  P  Y  T  F  T  P  T  T  Q  L  S  L  N  M  G  L  A  I
8301  ACAAATGAGCAATAATCTTAATATCACTAAACCTCCTGGGACTTCTACCGTATACATTCACACCAACAACAACTATCACTAAACATGGGACTTGCTAT  8400
        P  F  W  L  A  T  V  L  L  G  L  R  N  Q  P  T  A  A  L  G  H  L  L  P  E  G  T  P  T  L  L  I  P
8401  TCCATTCTGACTAGCAACAGTATTACTGGGACTGCGTAACCAACCCACTGCCGCGCTAGGACACCTTCTCCCAGAAGGAACACCAACCCTGCTAATCCCA  8500
        I  L  I  I  I  E  T  I  S  L  L  I  R  P  F  A  L  G  V  R  L  T  A  N  L  T  A  G  H  L  L  M  Q  L
8501  ATCCTAATTATTATTGAAACAATCAGCCTACTTATCCGCCCCTTCGCCCTAGGAGTACGACTAACAGCCAATCTCACAGCAGGCCACCTCCTAATACAAT  8600
        I  A  T  A  A  F  V  L  L  P  M  M  P  T  V  A  L  L  T  T  L  V  L  F  L  L  T  L  L  E  I  A  V
8601  TAATTGCTACCGCCGCCTTCGTACTCCTACCTATAATACCAACAGTAGCATTATTAACAACATTAGTCCTATTCCTCCTGACCCTGCTAGAAATTGCCGT  8700
                                                  COIII →
        A  M  I  Q  A  Y  V  F  V  L  L  L  S  L  Y  L  Q  E  N  V  M  A  H  Q  A  H  A  Y  H  M  V  D  P
8701  AGCAATAATCCAAGCCTACGTGTTTGTTCTATTACTAAGCCTCTATCTACAAGAAAATGTCTAATGGCCCACCAAGCACACGCATATCACATAGTTGACC  8800
        S  P  W  P  I  T  G  A  T  A  A  L  L  V  T  S  G  L  A  A  W  F  H  F  N  S  M  I  L  I  L  M  G
8801  CAAGCCCATGACCCATTACAGGGGCCACGGCCGCCCTACTTGTAACCTCAGGCCTAGCAGCGTGATTTCACTTCAACTCAATAATCTTAATTTTAATAGG  8900
        L  T  L  L  L  L  T  M  Y  Q  W  W  R  D  I  I  R  E  S  T  F  Q  G  H  H  T  L  P  V  Q  K  S  L
8901  ACTAACACTATTGCTACTAACTATGTATCAATGATGACGAGATATTATTGAGAAAGCACATTCCAAGGTCACCACACACTACCTGTACAAAAAAGGCCTA  9000
        R  Y  G  M  I  L  F  I  T  S  E  V  F  F  L  G  F  W  A  F  Y  H  S  S  L  A  P  T  P  E  L  G
9001  CGATATGGTATAATCCTGTTCATTACATCCGAAGTATTCTTCTTCCTAGGGTTCTTCTGAGCCTTTTACCATTCAAGTCTGGCACCCACTCCTGAACTCG  9100
        G  L  W  P  P  T  G  I  T  P  L  D  P  F  E  V  P  L  L  N  T  A  V  L  L  A  S  G  I  T  V  T  W
9101  GAGGACTCTGACCTCCCACTGGAATTACACCCCTAGATCCATTTGAAGTACCACTATTAAACACAGCAGTTCTACTAGCCTCGGGAATTACAGTCACATG  9200
        A  H  H  S  L  M  E  G  Q  R  K  E  A  I  Q  S  L  F  I  T  V  L  L  G  L  Y  F  T  A  L  Q  A  T
9201  AGCCCATCACAGCCTAATAGAGGGGCAACGAAAAGAGGCTATCCAATCACTATTTATCACAGTTCTGTTAGGACTATACTTTACAGCACTGCAAGCCACA  9300
        E  Y  Y  E  S  P  F  T  I  A  D  G  A  Y  G  S  T  F  F  V  A  T  G  F  H  G  L  H  V  I  I  G  S  T
9301  GAATACTACGAATCCCCATTTACAATCGCTGACGGAGCCTATGGCTCAACCTTTTTTGTAGCAACCGGATTCCACGGTCTACATGTCATTATTGGCTCTA  9400
        F  L  I  V  C  L  V  R  Q  T  Q  Y  H  F  T  S  N  H  H  F  G  F  E  A  A  A  W  Y  W  H  F  V  D
9401  CATTCCTAATCGTATGCCTAGTACGACAAACACAATACCACTTCACATCAAACCACCACTTTGGCTTTGAAGCAGCAGCATGATACTGACATTTCGTAGA  9500
        V  V  W  L  F  L  Y  V  S  I  Y  W  W  G  S  *  tRNA-Gly →
9501  CGTAGTCTGACTATTCTTATACGTATCAATCTACTGATGAGGCTCATAAACCCCTTTAGTACAAAATAATACAAATGACTTCCAATCATTTAATCTTGGT  9600
```

FIGURE 2.—*Continued*

NADH 3 →

```
        M  N  L  I  L  A  G  L  L  I  M  S  I  L  S  M  I  L  A  M  I  A  F  W  L  P  N
 9601 TATACCCCAAGAATGGGTAATGAACCTGATTCTAGCGGGCCTACTTATCATAAGCATCCTCTCTATAATTTTAGCTATAATCGCATTCTGACTACCAAAC  9700
        M  T  P  D  T  E  K  L  S  P  Y  E  C  G  F  D  P  L  G  S  A  R  L  P  F  S  L  R  F  F  L  V  A  I
 9701 ATGACCCCTGATACAGAAAAACTATCTCCCTACGAATGTGGCTTTGATCCTCTAGGATCCGCACGACTCCCATTCTCCCTACGATTTTTCCTAGTAGCAA  9800
        L  F  L  L  F  D  L  E  I  A  L  L  L  P  L  P  W  A  D  Q  L  T  N  P  T  L  A  L  T  W  T  T  S
 9801 TCCTATTCCTGCTATTTGACCTAGAAATTGCATTATTATTACCCCTACCCTGGGCAGACCAACTAACAAACCCAACACTTGCATTAACCTGGACAACAAG  9900
        I  I  A  L  L  T  L  G  L  I  H  E  W  T  Q  G  G  L  E  W  A  E    tRNA-Arg →       ___
 9901 CATCATCGCCCTACTAACACTAGGACTAATCCACGAATGAACTCAAGGAGGCCTCGAATGGGCAGAATAGGTAGTTAGTTTAAAAAAACCACTAATTTCG  10000
```

NADH 4L →

```
        M  T  P  V  Q  L  S  F  N  T  A  F  T  L  G  L  M  G  V  T  F
10001 ACTTAGTTAACTGTGGTGAAACCCCACAACTATCTTCATGACCCCAGTACAACTCAGCTTTAACACTGCATTCACACTAGGCTTAATAGGAGTAACATTC  10100
        H  R  A  H  L  L  S  A  L  L  C  L  E  G  M  M  L  S  L  Y  M  G  L  S  L  W  P  M  Q  L  E  S  T  T
10101 CACCGAGCCCATCTGCTATCAGCATTACTCTGCCTAGAAGGAATAATATTATCCCTGTATATAGGACTGTCCCTATGACCAATGCAACTAGAATCAACTA  10200
        Y  M  T  T  P  L  L  L  L  A  F  S  A  C  E  A  G  A  G  L  A  L  M  V  A  T  S  R  T  H  G  T  D
10201 CATACATAACCACACCACTACTACTACTCGCCTTCTCAGCCTGTGAGGCTGGAGCAGGCCTAGCCCTCATAGTAGCAACATCCCGCACACATGGTACGGA  10300
```

NADH 4 →

```
        M  L  K  V  L  M  P  T  I  M  L  I  L  T  T  W  L  T  K  P  A  W  L  W  P
        H  L  Q  N  L  N  L  L  Q  C  *
10301 CCACCTCCAAAACCTAAACTTACTACAATGCTAAAAGTTTTAATACCAACAATTATGCTTATCTTAACCACATGATTAACAAAACCTGCATGACTCTGAC  10400
        T  M  T  T  N  S  L  L  V  A  T  I  S  L  T  W  L  K  W  D  S  E  S  G  W  K  S  L  N  S  S  M  A
10401 CAACAATAACAACCAATAGCCTACTCGTAGCTACCATCAGCTTAACCTGACTAAAATGGGACTCAGAGTCAGGATGAAAATCTCTCAACAGCTCAATGGC  10500
        T  D  P  L  S  T  P  L  L  I  L  T  C  W  L  L  P  L  M  I  L  A  S  Q  N  H  M  F  M  E  P  L  N
10501 TACCGACCCCCTATCTACACCATTACTAATCCTCACATGCTGGCTTCTACCCCTCATAATTCTCGCAAGCCAAAACCACATGTTTATAGAACCACTAAAC  10600
        R  Q  R  S  F  I  S  L  L  I  S  L  Q  T  F  L  I  M  A  F  G  A  T  E  I  I  L  F  Y  I  M  F  E  A
10601 CGCCAACGATCATTCATCTCCCTACTCATCTCCCTACAAACATTCCTAATTATAGCATTTGGTGCCACTGAAATCATCCTATTTTTACATTATATTTGAAG  10700
        T  L  I  P  T  L  I  I  I  T  R  W  G  N  Q  T  E  R  L  N  A  G  T  Y  F  L  F  Y  T  V  M  G  S
10701 CAACCCCTAATCCCAACACTAATTATTATTACCCGATGGGGTAATCAAACAGAGCGACTAAACGCAGGAACATACTTTTTATTTTATACAGTAATAGGGTC  10800
        L  P  L  L  V  A  L  L  M  T  Q  N  N  L  G  T  L  S  M  P  L  I  Q  H  M  Y  Q  M  K  L  H  T  H
10801 ACTACCACTATTAGTTGCACTTTTAATAACACAAAATAACCTTGGTACCCTATCAATACCGCTCATCCAACACATATACCAAATAAAACTTCATACACAT  10900
        G  D  M  M  W  W  T  A  C  L  L  A  F  L  V  K  M  P  L  Y  G  V  H  L  W  L  P  K  A  H  V  E  A  P
10901 GGAGACATGATATGATGAACAGCCTGCCTATTAGCCTTCTTAGTAAAAATACCACTATACGGAGTCCACCTTTGACTCCCAAAAGCCCATGTAGAAGCCC  11000
        I  A  G  S  M  V  L  A  A  V  L  L  K  L  G  G  Y  G  M  M  R  L  I  M  M  L  A  P  M  T  K  T  L
11001 CAATTGCAGGATCAATAGTACTAGCCGCCGTCCTACTAAAACTAGGAGGATACGGAATAATACGACTAATCATAATATTAGCTCCAATAACAAAAACCCT  11100
        A  Y  P  F  I  I  L  A  L  W  G  I  I  M  T  G  S  I  C  L  R  Q  T  D  L  K  S  L  I  A  Y  S  S
11101 AGCCTATCCATTCATCATCCTCGCCCTATGAGGAATCATTATAACTGGATCAATCTGCTTACGACAAACAGACCTAAAATCCCTAATCGCCTACTCATCA  11200
        V  G  H  M  G  L  V  A  A  G  I  L  T  Q  T  P  W  G  F  T  G  A  T  V  L  M  I  A  H  G  L  T  S  S
11201 GTAGGCCACATAGGACTAGTGGCAGCAGGTATCCTAACACAAACACCATGAGGCTTTACAGGAGCTACTGTTCTAATAATTGCTCACGGTCTTACATCCT  11300
        A  L  F  C  L  A  N  T  N  Y  E  R  T  H  S  R  T  M  I  L  A  R  G  M  Q  V  I  L  P  L  M  T  F
11301 CAGCCCTATTCTGTCTAGCAAACACAAACTATGAACGAACCCATAGCCGAACCATGATCCTAGCACGAGGAATGCAAGTTATCCTCCCACTCATGACATT  11400
        W  W  L  M  M  N  L  A  N  L  A  L  P  P  S  T  N  L  M  G  E  L  L  I  I  T  M  T  F  N  W  S  N
11401 CTGATGACTTATAATAAATTTAGCTAACTTAGCCTTACCCCCATCCACTAATCTAATAGGAGAACTACTTATTATTACAATAACCTTCAACTGATCAAAC  11500
        W  T  L  T  M  T  G  L  G  M  L  I  T  A  I  Y  S  L  H  M  F  L  T  T  Q  R  G  L  M  T  N  H  I  I
11501 TGGACACTAACTATAACAGGACTGGGCATACTAATCACAGCTATCTACTCATTACACATGTTCCTCACAACACAACGAGGCCTGATAACAAACCACATTA  11600
        S  Y    tRNA-His →                ___                              tRNA-Ser(AGY) →
11601 TCTCAATTGAACCCTCCCACACCCGAGAACACCTACTAATAACAATACATGCCCTCCCAATACTATTGCTAATCCTTAAACCAGAACTGATCTGAGGCTG  11700
11701 ATCCTACTGTAAATATAGTTTGAAAAAAACATTAGACTGTGGCTCTAAAAATAAGAGTTAAATCCTCTTTATTAACCGAGAGGGGCTAGAGCACTACGAA  11800
              ___                        tRNA-Leu(CUN) →                                    ___
11801 CTGCTAATTCTGTAGTACCATGGTTCAATCCCATGGCCCACTCAGCCCCTAAAGGATAATAGCTCATCCATTGGTCTTAGGAGCCAAAAACTCTTGGTGC  11900
```

NADH 5 →

```
        M  Y  T  T  L  I  F  N  S  T  L  M  T  M  F  T  I  L  T  A  P  I  L  T  T  L  N  P
11901 AACTCCAAGTAGCGGCCATGTACACAACATTAATTTTTAACTCAACACTTATAACCATATTTACCATCCTAACAGCCCCAATCTTAACCACCCTTAACCC  12000
        I  K  P  N  K  K  W  T  E  L  W  V  K  T  A  V  Q  L  A  F  F  T  S  L  I  P  L  F  I  Y  L  D  Q
12001 TATCAAACCCAACAAAAGTGAACAGAATTATGAGTAAAAACCGCCGTACAGCTAGCCTTCTTCACAAGTCTAATCCCGCTCTTCATCTATCTTGACCAA  12100
        G  I  E  T  I  T  T  N  W  Q  W  M  N  T  N  F  N  I  N  I  S  F  K  F  D  Q  Y  S  I  V  F  I  P
12101 GGAATCGAGACCATTACAACAAACTGACAATGAATAAACACCAACACATTTAACATTAACATCAGCTTTAAATTTGACCAATACTCAATTGTATTCATCC  12200
        I  A  L  Y  V  T  W  S  I  L  E  F  A  N  W  Y  M  H  Q  D  P  K  M  N  Q  F  F  K  Y  L  L  L  F
12201 CAATCGCACTGTATGTCACATGATCAATTCTAGAATTCGCCAACTGATACATGCACCAAGACCCAAAAATAAACCAATTTTTCAAATACCTGCTCCTATT  12300
        L  I  T  M  M  I  L  V  T  A  N  N  M  F  Q  L  F  I  G  W  E  G  V  G  I  M  S  F  L  L  I  G  W
12301 CCTAATCACCATAATAATCTTAGTTACAGCAAACAACATGTTCCAACTATTTATCGGCTGAGAAGGAGTAGGAATTATATCATTCCTATTAATTGGCTGA  12400
        N  L  N  T  W  E  M  Q  Q  M  F  I  M  S  Y  N  T  D  M  T  I  P  L  L  G  L  T  L  A  A  A  G  K
12401 TGACATGGCCGAGCTAACGCCAATACCGCAGCCCTACAAGCAGTCATCTACAATCGAGTAGGAGACATTGGGTTAACCCTAAGCATGGTCTGATTTGCAA  12500
12501 TCAACCTAAATACATGAGAAATACAACAAATATTTATCATATCCTACAACACCGACATAACCATCCCCCTACTAGGCCTAACTCTGGCAGCAGCAGGAAA  12600
```

FIGURE 2. — *Continued*

```
          S  A  Q  F  G  L  H  P  W  L  P  A  A  M  E  G  P  T  P  V  S  A  L  L  H  S  S  T  M  V  V  A  G
12601 ATCAGCCCAATTTGGATTACACCCATGACTACCAGCAGCTATAGAAGGTCCAACACCGGTCTCTGCCCTACTACACTCAAGTACCATAGTGGTTGCCGGA  12700
          I  F  L  L  I  R  L  H  P  L  M  D  N  N  K  L  I  L  T  T  C  L  C  L  G  A  L  T  T  L  F  T  A  A
12701 ATCTTCCTGCTTATCCGACTACACCCCCTCATAGACAATAATAAACTAATCCTCACTACCTGCCTCTGCCTAGGAGCACTAACCACCCTATTTACTGCCG  12800
          C  A  L  T  Q  N  D  I  K  K  I  I  A  F  S  T  S  S  Q  L  G  L  M  M  V  A  I  G  L  N  Q  P  Q
12801 CATGCGCACTCACCCAAAACGATATTAAAAAAATTATTGCATTTTCAACATCAAGTCAACTAGGACTAATAATGGTAGCAATTGGACTAAATCAACCACA  12900
          L  A  F  L  H  I  C  T  H  A  F  F  K  A  M  L  F  L  C  S  G  S  I  I  H  S  L  N  D  E  Q  D  I
12901 ACTAGCATTCCTCCACATCTGCACCCACGCTTTCTTTAAAGCAATGCTGTTCCTATGCTCCGGATCAATTATTCACAGTCTAAATGATGAACAAGATATC  13000
          R  K  M  G  G  L  H  N  M  L  P  T  T  S  S  C  T  M  V  S  S  M  A  L  T  G  M  P  F  L  A  G  F  F
13001 CGAAAAATAGGTGGCTTACACAACATGCTCCCAACAACAAGCTCCTGTACAATAGTTAGCAGCATAGCCCTAACAGGAATGCCATTCCTAGCAGGCTTCT  13100
          S  K  D  A  I  I  E  S  L  N  S  S  H  L  N  A  W  A  L  T  T  T  L  M  A  T  S  F  T  A  V  Y  S
13101 TCTCAAAAGACGCAATCATCGAATCACTAAACTCCTCTCACCTAAACGCCTGAGCCCTAACTACTACACTAATAGCCACATCATTCACCGCAGTCTATAG  13200
          L  R  I  I  Y  L  V  M  M  N  Y  P  R  F  Q  S  L  S  P  I  D  E  N  Y  T  K  T  R  N  P  I  K  R
13201 CCTCCGAATTATTTACCTTGTAATGATAAACTATCCACGATTTCAAAGCCTGTCCCCCATCGATGAAAATTATACAAAAACACGTAACCCAATTAAACGA  13300
          L  A  W  G  S  V  I  A  G  M  M  I  F  L  N  I  L  P  T  K  S  Q  T  M  T  M  P  T  H  M  K  T  A  A
13301 CTAGCATGAGGAAGCGTAATCGCAGGAATAATGATCTTCCTAAACATCCTACCAACTAAATCACAAACAATGACCATACCCACCCATATAAAAACAGCCG  13400
          I  M  V  T  I  L  G  V  L  T  A  M  E  L  T  K  L  T  S  T  Q  L  K  I  T  P  N  V  N  M  H  N  S
13401 CAATCATAGTTACAATCCTAGGAGTCCTCACCGCTATAGAACTAACAAAACTTACAAGCACACAACTAAAAATCACCCCAAACGTCAACATACACAACTC  13500
          T  W  F  E  K  I  G  P  K  G  I  L  A  L  Q  I  P  T  T  K  M  I  N  N  S  Q  Q  G  L  I  K  T  Y  L
13501 ATCCAACATACTAGGATACTTCCCAAACATCATTCACCGACTTCTACCACAAACAAACTTATATCTAGGACAAAAAATAGCAACACACCTAACAGATCAA  13600
          T  L  F  F  L  T  T  V  L  F  T  T  M  T  M  I  *         *  V  A  R  L  A  G  R  S  L  G  R  T  I  E  L  V
13601 ACATGATTTGAAAAAATCGGGCCAAAAGGAATTTTAGCCCTACAAATCCCTACAACTAAAATAATCAACAACTCGCAACAAGGACTAATCAAAACATACC  13700
          V  F  L  A  L  L  L  V  W  G  T  V  L  L  L  F  G  G  S  S  Y  V  M  P  V  G  S  V  D  G  S  L  V
13701 TGACCCTATTTTTTCTAACAACAGTTCTATTTACTACAATAACCATAATCTAACTGCTCGAAGCGCACCACGGGACAACCCACGAGTAATCTCCAAAACC  13800
          S  M  E  G  F  V  D  L  S  M  W  Y  F  D  Y  W  W  G  Y  C  Y  Y  G  V  L  L  F  G  F  L  Y  V  L
13801 ACAAACAAAGCCAACAATAACACCCACCCAGTCACCAACAACAAAAATCCCCCAGAAGAATAAACTATGGGCACACCACTGACATCCCCTGACAAACTG  13900
          Y  L  F  V  S  W  S  G  W  S  E  P  Y  P  E  A  A  L  A  A  S  Y  A  F  V  V  L  M  G  G  L  Y  I
13901 ATATCTCACCAAAAACATCCAATGATATTCAATAAAAATCATACCACCACCCATAACAATAATACCCAACAAGTAAAAAGCCAAACAAATAAACCAAAAT  14000
          L  F  L  V  L  S  L  F  S  G  G  Y  G  V  L  F  G  C  G  V  V  A  S  F  V  L  G  L  A  A  F  Y  P
14001 ATAAAGAAAAACCGACCAACTCCCCCAAGACTCAGGATAAGGCTCTGCTGCTAAAGCAGCAGAATAAGCAAAAACAACCAATATCCCCCCCAAATAAATC  14100
          A  P  N  S  A  V  A  M  L  G  V  V  L  G  M  L  V  V  F  V  F  Y  I  M
14101 AAAAACAGTACTAAGGACAAAAAAGAGCCACCGTACCCAACCAAAAAGCCACAGCCAACAACTGCAGAAAATACTAACCCAAGAGCAGCAAAATAAGGAG  14200
                                                          ←NADH 6
14201 CTGGATTAGACGCAACTGCCATCAAACCAACCACGAGTCCCATCAACACAACAAAAACAAAATAAATCATAATTCTTGCCCGGACTTTAACCAGAACCAA  14300
                                                               Cyt b →
          ___                              ←tRNA-Glu  M  T  N  I  R  K  T  H  P  L  I  K  I  I  N  E  T  I  I  D
14301 TGACTTGAAAAACCACCGTTGTCAATTCAACTACAAAAACCCATGACAAACATCCGAAAGACACACCCGCTAATTAAAATTATCAACGAAACCATCATCG  14400
          L  P  T  P  S  N  I  S  I  W  W  N  F  G  S  L  L  G  I  C  L  I  T  Q  I  V  T  G  L  F  L  A  M
14401 ACCTCCCCACACCATCAAACATCTCAATCTGATGAAATTTTGGGTCACTACTAGGAATTTGTTTAATTACACAAATCGTAACAGGCCTATTCTTAGCAAT  14500
          H  Y  T  A  D  I  T  T  A  F  S  S  V  A  H  I  C  R  D  V  N  Y  G  W  L  I  R  S  T  H  A  N  G
14501 ACACTACACAGCTGACATTACAACAGCATTCTCATCAGTAGCCCACATCTGCCGAGATGTAAACTATGGATGACTAATCCGAAGTACCCATGCCAACGGA  14600
          A  S  L  F  F  I  C  I  Y  L  H  V  A  R  G  L  Y  Y  G  S  Y  L  Q  K  E  T  W  N  I  G  V  I  L  L
14601 GCCTCTCTATTCTTCATCTGCATCTACCTACATGTAGCACGTGGACTCTACTATGGGTCATACTTACAAAAAGAAACCTGAAACATCGGAGTTATCCTCC  14700
          M  L  V  M  I  T  A  F  V  G  Y  V  L  P  W  G  Q  M  S  F  W  G  A  T  V  I  T  N  L  L  S  A  V
14701 TCATGCTAGTTATGATTACTGCCTTCGTGGGATATGTCCTCCCCTGAGGCCAAATATCATTCTGAGGGGCAACCGTCATCACAAACCTCCTGTCAGCAGT  14800
          P  Y  I  G  D  T  L  V  Q  W  I  W  G  G  F  S  V  D  N  A  T  L  T  R  F  F  A  F  H  F  L  L  P
14801 ACCCTATATTGGAGATACACTAGTTCAATGAATCTGAGGAGGTTTTTCCGTAGACAACGCCACACTCACGCGATTCTTTGCCTTCCACTTCCTCCTGCCA  14900
          F  V  I  S  G  A  S  I  I  H  L  L  F  L  H  E  T  G  S  N  N  P  T  G  L  N  S  D  A  D  K  V  T  F
14901 TTCGTAATCTCAGGAGCCTCGATTATTCACCTGCTCTTCCTTCATGAGACAGGATCCAACAACCCAACTGGCCTTAACTCTGACGCAGATAAAGTAACCT  15000
          H  P  Y  F  S  Y  K  D  M  L  G  F  L  I  T  L  T  T  L  A  F  L  T  L  F  T  P  N  L  L  G  D  P
15001 TCCACCCGTACTTCTCGTACAAAGACATGTTAGGATTCCTAATTACACTAACAACACTAGCATTCCTAACCCTATTCACTCCAAACCTGTTAGGAGACCC  15100
          E  N  F  T  P  A  N  P  L  T  T  P  P  H  I  K  P  E  W  Y  F  L  F  A  Y  A  I  L  R  S  I  P  N
15101 CGAAAACTTCACACCAGCAAACCCATTAACCACCCCGCCACACATCAAACCAGAATGATACTTCCTATTCGCCTATGCAATTCTACGATCCATCCCCAAC  15200
          K  L  G  G  V  L  A  L  V  S  S  I  L  V  L  L  L  V  P  I  L  H  T  S  K  Q  R  G  N  T  F  R  P  I
15201 AAACTAGGAGGAGTTCTAGCTCTAGTCTCCTCTATTCTAGTACTACTACTAGTGCCAATCTTACACACCTCAAAACAACGAGGAAACACCTTCCGCCCAA  15300
          T  Q  M  L  F  W  A  L  V  A  D  M  L  I  L  T  W  I  G  G  Q  P  V  E  Y  P  F  M  T  I  G  Q  I
15301 TTACCCAAATACTATTCTGAGCCCTTGTAGCAGACATGCTAATCCTAACCTGAATCGGAGGCCAACCAGTAGAATATCCATTCATGACAATCGGACAAAT  15400
          A  S  I  T  Y  F  S  L  F  L  I  L  I  P  M  T  G  W  L  E  N  K  A  M  N  W  N     tRNA-Thr →
15401 TGCTTCAATCACCTACTTCAGCCTATTCCTCATCCTAATCCCAATAACTGGATGACTAGAAAACAAGGCCATAAACTGAAACTAGCCTTAGTAGCTTAAA  15500
          ___
15501 CCCAAAGCACCGGCCTTGTAAGCCGGAGATTGAGAACTGACCATTCTCCTAAAGCTTTCAGAGAAAAAAGGATCAAACTTTTACCCTTAGCTCCCAAAGC  15600
                              ←tRNA-ProControl region→
15601 TAAAATTCTAATTAAACTATTCTCTGTTCCCAAGCTCTGCCCATCCACTCAATAACCCCCCTCCTTCTAGTACGTACTATCTCTATGTATATCGAACATT  15700
```

FIGURE 2.— *Continued*

```
15701  CGTTACTCTCAAGTACATTATACATTACATGTTAATTTACCATTAGAACTGTCACACCACATTTAGGGCGAGAAACCAGCAATACTTGCATAAATGATTA  15800

15801  ATTAATTTACGATGTGGTTAGACACTTATTTCTTAATCTAAACTTGGCTTTATCCATTACTGGCCACTGGTACTGTGCGATGGAGAATAATAGAAAGATT  15900

15901  TATATGATTATAAATTATCTATTACTGGCATCTGGTTTTGGGTTTAGTGAGGGGAAGGGCTTTTTAACCCGTAACTCAGTATCACTTTACTATACTGGCC  16000

16001  AGCCGTTATTTTTTTGGGGGGGGGGAACGTGAAGTTAGACAGCACTTAACTTGCGTCAAACGTCGTTTATGATTGGACTTTTAGTCGACACTCAAGTACT  16100
```

```
                                            _____Repeat 1_____/_____Repeat 2____/____Repeat 3_____
16101  TTTGGATCTATGACAAAGGATAATCAGTTAATGATAGATAGACATATAATGAATGATAGATAGACATATAATGAATGATAAATAGACATATAATGAATGA  16200

       ___/_____Repeat 4_____
16201  TAGTTAGATATATAGATTAATGTAAGAAAGATAATTGAACCCATGACAGAGGACATACTTTTAATGATTTCAGGACATAAACACCATGCACATCAGATTG  16300

16301  TATCAACTAAACCCCTCCCAACCCCTAAAACCCAGGACTCGCTAAACACATCAACCCATATATTTTTTCAGTATATACATTGTTATACATCAAATAAATA  16400

16401  ATGTAAT  16407
```

FIGURE 2.—*Continued*

phological and molecular characters that could have evolved at the time of this rapid radiation have been obliterated by subsequent changes and back mutations during the last 400 million years. Based on the currently available data, it had remained difficult to distinguish between competing phylogenetic hypotheses, and larger molecular data sets are needed to address this important phylogenetic issue.

Mitochondrial DNA has been widely and successfully used in the past to infer phylogenetic relationships among many different species and might therefore be expected to be a suitable candidate for the evolutionary question at hand. However, several authors (*e.g.*, CUM-MINGS *et al.* 1995; RUSSO *et al.* 1996; ZARDOYA and MEYER 1996c) recently pointed out the high risk of not recovering the exact evolutionary relationships of the taxa under study when individual mitochondrial genes are analyzed and have suggested that larger or more heterogeneous data sets, *i.e.*, complete mitochondrial genomes, are needed to confidently resolve some phylogenetic questions. To address the question of the relationships among living sarcopterygians and to study the evolution of the mitochondrial genome in vertebrates, we previously sequenced the complete mitochondrial genome of an African lungfish (ZARDOYA and MEYER 1996a). Here, we present the complete nucleotide sequence of the mitochondrial genome of the coelacanth, *L. chalumnae*.

MATERIALS AND METHODS

DNA extraction, PCR amplification, cloning and sequencing: The scarcity of coelacanth samples (BRUTON and COU-TOUVIDIS 1991; FRICKE 1992; FRICKE *et al.* 1995) and their generally poor quality did not allow us to isolate mitochondria directly from tissue and to obtain intact mitochondrial genomes. Therefore, a total DNA extraction was performed (TOWNER 1991) and the isolated DNA was cleaned through a Sephadex G-50 column. A combination of 24 sets of versatile primers was designed (Table 1) based on highly conserved vertebrate mtDNA regions. They were used to amplify via PCR, contiguous and overlapping fragments (averaging ~800 bp) which covered the entire coelacanth mtDNA molecule (Figure 1). Some of these primers are expected to successfully amplify mitochondrial DNA fragments in other related vertebrate species (*e.g.*, the primer pairs 19–20 and 27–28 of Table

1 have successfully amplified in fishes and amphibians, unpublished data). Thirty-five cycles of PCR (denaturing at 94° for 60 sec, annealing at 45–50° for 60 sec and extending at 72° for 60–105 sec) were performed in 25-μl reactions containing 67 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.4 mM of each dNTP, 2.5 μM of each primer, template DNA (10–100 ng), and AmpliTaq DNA polymerase (1 unit, Perkin-Elmer-Cetus).

PCR products were cloned using the pGEM-T vector (Promega) and sequenced using the FS-Taq Dye Deoxy Terminator cycle-sequencing kit (Applied Biosystems Inc.) with an automated DNA sequencer (Applied Biosystems 373A) as previously described (ZARDOYA and MEYER 1996a). DNA sequences were obtained using both M13 universal sequencing primers, and three specific oligonucleotide primers for those clones containing inserts >1000 bp (see Table 1). Typically, only one clone per PCR product was sequenced. No differences were found in any of the overlapping sequences between contiguous clones. The fidelity of Taq polymerase is ~5 × 10^{-6} errors per nucleotide incorporated per cycle (GELFAND and WHITE 1990) (one error in 16,400 nucleotides after 35 cycles). However, recently XU and ARNASON (1996) have reported that the sequence of the gorilla mitochondrial genome obtained by direct isolation from mitochondria differed outside the control region by 49 nucleotides with respect to the corresponding mitochondrial genome obtained partially by PCR (HORAI *et al.*, 1995). Obviously, not all these differences are due to Taq polymerase error during amplification but also are due to sequencing errors, so we expect in a more realistic estimate that our sequence might differ maximally (outside the control region) by <20 nucleotides with respect to the hypothetical sequence from a genome obtained by direct isolation from mitochondria. Based on these calculations, we can submit that the sequence reported is representative of the coelacanth mitochondrial genome and that, although we recommend traditional methods for obtaining mitochondrial genomes if possible, our PCR approach to obtain the complete mitochondrial genome sequence will be useful for other rare or endangered species.

Molecular and phylogenetic analyses: Sequence data were analyzed with the GCG program package (DEVEREUX *et al.* 1984), MacClade version 3.06 (MADDISON and MADDISON 1992) and PAUP* version d54 (SWOFFORD 1997). DNA sequences were aligned using CLUSTAL W (THOMPSON *et al.* 1994) followed by refinement by eye. Ambiguous alignments, mainly in 5' and 3' ends of protein-coding genes, in tRNA gene sequences corresponding to the tRNA DHU and TψC arms, and in several highly variable regions of the rRNA genes, were excluded from the phylogenetic analyses (aligned sequences are available from the authors upon request). Third codon positions were also excluded from all phylogenetic analyses. Transitions and transversions were given equal weight

TABLE 2

Localization of features in the mitochondrial genome of the coelacanth

Feature	From	To	Size (bp)	Codon Start	Stop
tRNA-Phe	1	69	69		
12S rRNA	70	1052	983		
tRNA-Val	1053	1119	67		
16S rRNA	1120	2784	1665		
tRNA-Leu (UUR)	2785	2860	76		
NADH 1	2861	3832	972	ATG	TAA
tRNA-Ile	3833	3906	74		
tRNA-Gln	3974	3904	69 (L)		
tRNA-Met	3974	4043	70		
NADH 2	4044	5090	1047	ATG	TAA
tRNA-Trp	5093	5164	72		
tRNA-Ala	5235	5167	67 (L)		
tRNA-Asn	5310	5238	71 (L)		
tRNA-Cys	5403	5338	65 (L)		
tRNA-Tyr	5475	5404	69 (L)		
CO I	5477	7024	1548	GTG	TAA
tRNA-Ser (UCN)	7100	7030	69 (L)		
tRNA-Asp	7105	7174	70		
CO II	7180	7870	691	ATG	T--
tRNA-Lys	7871	7943	73		
ATPase 8	7944	8111	168	ATG	TAA
ATPase 6	8102	8763	662	ATG	TA-
CO III	8764	9549	786	ATG	TAA
tRNA-Gly	9550	9619	70		
NADH 3	9620	9968	349	ATG	T--
tRNA-Arg	9969	10,036	68		
NADH 4L	10,038	10,334	297	ATG	TAA
NADH 4	10,328	11,708	1381	ATG	T--
tRNA-His	11,709	11,777	69		
tRNA-Ser (AGY)	11,778	11,844	67		
tRNA-Leu (CUN)	11,845	11,917	73		
NADH 5	11,918	13,753	1836	ATG	TAA
NADH 6	14,270	13,750	519 (L)	ATG	AGA
tRNA-Glu	14,340	14,271	70 (L)		
Cyt b	14,343	15,484	1142	ATG	TA-
tRNA-Thr	15,485	15,556	72		
tRNA-Pro	15,626	15,558	67 (L)		
Control region	15,627	16,407	781		

(for exceptions see below). Overlapping positions (two open reading frames) in several genes (ATPase8/ATPase6, ND4L/ND4, ND5/ND6) were duplicated in the analyses. Gaps resulting from the alignment were treated as missing data. The control region was also excluded from the analyses due to its fast rate of evolution which prevented reliable alignment and made it not appropriate for this phylogenetic question.

Three data sets (protein, rRNA, and tRNA coding genes) were combined and subjected to the maximum-parsimony (MP) method (PAUP version 3.1.1, SWOFFORD 1993; and PAUP* version D54; SWOFFORD 1997), using heuristic searches (TBR branch swapping; MULPARS option in effect) with 10 random stepwise additions of taxa to find the most parsimonious trees. Neighbor-joining (NJ) (SAITOU and NEI 1987) (based on Kimura-corrected distance matrices, jumble option in effect) and maximum likelihood (ML) (transversions were given double the weight of transitions; empirical base frequencies and five random stepwise addition of taxa were used) analyses of the sequences were performed with

PHYLIP (version 3.55) (F84 model, FELSENSTEIN 1989), MOLPHY version 2.2 (ADACHI and HASEGAWA 1992), and PAUP* version d54 (SWOFFORD 1997). Robustness of the inferred trees was tested by bootstrapping (FELSENSTEIN 1985) (as implemented in PAUP version 3.1.1., PAUP*, and PHYLIP with 100 pseudoreplications each).

Statistical methods: The statistical confidence of MP analyses was evaluated by calculating the standard deviation of the difference in number of steps between the resulting most parsimonious trees and the two alternative hypotheses using the method of TEMPLETON (1983) as implemented in PHYLIP. Similarly, the statistical confidence of the resulting best tree of the ML analysis with respect to competing hypotheses was assessed by calculating the standard deviation of the difference in log-likelihood between the resulting best tree and the alternative hypotheses using the formula of KISHINO and HASEGAWA (1989) as implemented in PHYLIP, PAUP*, and MOLPHY. In both cases, competing trees were declared significantly different if the difference in number of steps or

TABLE 3

Base composition of vertebrate mitochondrial genomes

	A	C	G	T
Proteins				
Tetrapods				
1	30.7	25	21.2	23.1
2	19.3	27.1	12.4	41.2
3	41	31.9	4.8	22.2
Total	30.3	28	12.8	28.9
Lungfish				
1	27.6	25.4	23.6	23.4
2	18.4	27.2	13.3	41.1
3	34.7	28.5	8.4	28.4
Total	26.9	27	15.1	31
Coelacanth				
1	30.2	26	23.2	20.6
2	18.4	27.3	13.3	41
3	47.7	28.3	6.9	17.1
Total	32.4	27.1	14.5	26
Teleosts				
1	26.2	26.6	26.2	21
2	18.5	27.4	13.8	40.3
3	37.8	33.2	8.1	20.9
Total	27.5	29.1	16	27.4
Bichir				
1	29.9	23.7	23	23.4
2	18.9	27	12.8	41.3
3	42.7	27.8	4.4	25.1
Total	30.5	26.2	13.4	29.9
Lamprey				
1	30.4	22.9	22.6	24.1
2	19	26.5	12.9	41.6
3	41.3	21.5	3.8	33.4
Total	30.2	23.7	13.1	33
tRNAs				
Tetrapods	31.9	18.6	19.8	29.7
Lungfish	28.3	20.9	23.6	27.2
Coelacanth	30	20.8	21.8	27.4
Teleosts	28.2	21.3	23.5	27
Bichir	30.7	18.8	20.8	29.7
Lamprey	30.4	18.8	20.4	30.4
rRNAs				
Tetrapods	35.4	23.7	17.7	23.2
Lungfish	33	22.7	20.1	24.2
Coelacanth	36.7	24.9	18.6	19.8
Teleosts	33.4	25.4	21.5	19.7
Bichir	34.5	22	19.9	23.6
Lamprey	36.1	22.6	17.7	23.6

Tetrapods: human, Blue whale, opossum, chicken, and frog; teleosts: Rainbow trout, carp, and loach.

Pag-likelihoods where found to be >1.96 times the standard deviations (FELSENSTEIN 1989).

The complete mtDNA sequence of the coelacanth has been deposited at the EMBL/GenBank data libraries under accession no. U82228.

RESULTS AND DISCUSSION

Genome organization and base composition: The complete L-strand nucleotide sequence of the coela-

canth mitochondrial genome (16,407 bp) is depicted in Figure 2. The organization of the coelacanth mitochondrial genome conforms to the consensus vertebrate mitochondrial gene order (Figures 1 and 2, Table 2). As in other vertebrates, two rRNAs, 22 tRNAs and 13 proteins are encoded by the coelacanth mitochondrial genome. The overall base composition of the L strand is A: 34%; T: 24%; C: 27%; and G: 15%. As in other vertebrate mitochondrial genomes, guanine is the rarest nucleotide whereas adenine is the most frequent (MEYER 1993). A more detailed analysis of the base composition was performed by considering the rRNA, the tRNA and the protein coding genes separately (Table 3). In coelacanth protein-coding genes, there is an anti-G bias in third codon positions and pyrimidines are overrepresented in second codon positions, as was noted before for other vertebrate mitochondrial genomes (NAYLOR et al. 1995). Coelacanth tRNAs are A+T rich (57%) whereas the rRNAs have a high adenine content (Table 3). The noncoding intergenic spacer regions (22 bp), which are likely not subjected to strong selection, showed an A+C bias (73%). This indicates that, as is the case in other vertebrate mitochondrial genomes, an asymmetrical directional mutation pressure may be operating in the coelacanth genome (JERMIIN et al. 1995).

Noncoding sequences: The control region in the coelacanth mitochondrial genome is 781 bp long and 65% A+T rich (Figure 2). The coelacanth mitochondrial control region is characterized by the presence of four 22-bp tandem repeats in the right domain, close to the 3' end. Of these, three are perfect repeats whereas one is imperfect (six nucleotides are different). Even with the presence of these repeats, the coelacanth mitochondrial control region is, so far, the shortest among vertebrates (excluding the atypical control region of the lamprey) (LEE and KOCHER 1995). Neither conserved sequence blocks (CSBs, WALBERG and CLAYTON 1981; SOUTHERN et al. 1988; DILLON and WRIGHT 1993) nor termination associated sequences (TASs, DODA et al. 1981; FORAN et al., 1988), which are commonly found in other vertebrate mitochondrial control regions, could unambiguously be identified in the coelacanth mitochondrial control region. Putative CSB-II and -III motifs, sharing limited sequence similarity to the human and mouse consensus sequence (WALBERG and CLAYTON 1981), were tentatively identified at positions 16,308 and 16,324, respectively (see Figure 2).

The origin of light strand replication (O_L) of the coelacanth mitochondrial genome is, as in most vertebrates, located in a cluster of five tRNA genes (WANCY region) (but see SEUTIN et al. 1994) (Figure 2). This region is 26 nucleotides in length and has the potential to fold into a stem-loop secondary structure. The folding of the O_L does not require the use of part of the adjacent tRNA[CYS] as has been described for the lungfish mitochondrial genome (ZARDOYA and MEYER 1996a).

TABLE 4

Amino acid composition of vertebrate mitochondrial proteins

	Tetrapods	Lungfish	Coelacanth	Teleost	Bichir	Lamprey	Average
Ala	260 (GCC)	312 (GCC)	286 (GCA)	344 (GCC)	307 (GCC)	297 (GCC)	301 (GCC)
Arg	67 (CGA)	71 (CGA)	73 (CGA)	78 (CGA)	74 (CGA)	68 (CGA)	72 (CGA)
Asn	151 (AAC)	143 (AAY)	147 (AAC)	117 (AAC)	147 (AAC)	143 (AAT)	141 (AAC)
Asp	67 (GAC)	71 (GAC)	71 (GAC)	76 (GAC)	73 (GAC)	67 (GAC)	71 (GAC)
Cys	27 (TGC)	28 (TGY)	25 (TGC)	26 (TGC)	29 (TGC)	40 (TGT)	29 (TGC)
Gln	93 (CAA)	100 (CAA)	103 (CAA)	100 (CAA)	98 (CAA)	98 (CAA)	99 (CAA)
Glu	92 (GAA)	93 (GAA)	102 (GAA)	102 (GAA)	94 (GAA)	93 (GAA)	96 (GAA)
Gly	215 (GGA)	246 (GGA)	239 (GGA)	249 (GGA)	226 (GGA)	213 (GGA)	231 (GGA)
His	100 (CAC)	93 (CAC)	103 (CAC)	105 (CAC)	98 (CAY)	106 (CAC)	101 (CAC)
Ile	333 (ATY)	318 (ATT)	285 (ATC)	281 (ATY)	356 (ATT)	334 (ATT)	318 (ATT)
Leu	626 (CTA)	650 (CTW)	622 (CTA)	637 (CTA)	611 (CTA)	610 (YTA)	626 (CTA)
Lys	93 (AAA)	72 (AAA)	84 (AAA)	70 (AAA)	82 (AAA)	99 (AAA)	83 (AAA)
Met	210 (ATA)	176 (ATA)	224 (ATA)	164 (ATA)	212 (ATA)	217 (ATA)	200 (ATA)
Phe	226 (TTC)	242 (TTT)	214 (TTC)	225 (TTY)	214 (TTY)	235 (TTT)	223 (TTY)
Pro	211 (CCM)	212 (CCA)	206 (CCA)	209 (CCM)	203 (CCA)	202 (CCA)	207 (CCA)
Ser	288 (TCM)	258 (TCM)	226 (TCA)	235 (TCM)	256 (TCA)	263 (TCW)	254 (TCM)
Thr	326 (ACM)	295 (ACA)	359 (ACA)	305 (ACC)	299 (ACA)	306 (ACA)	315 (ACA)
Trp	107 (TGA)	116 (TGA)	121 (TGA)	120 (TGA)	116 (TGA)	109 (TGA)	114 (TGA)
Tyr	125 (TAY)	120 (TAY)	117 (TAY)	114 (TAC)	118 (TAY)	108 (TAT)	117 (TAY)
Val	169 (GTA)	172 (GTA)	183 (GTA)	222 (GTA)	174 (GTA)	192 (GTW)	185 (GTA)
Stop	13 (TAA)	13 (TAA)	13 (TAA)	13 (TAA)	13 (TAA)	13 (AGA/TAA)	13 (TAA)
Total	3799	3801	3803	3792	3800	3813	3801

Tetrapods: human, Blue whale, opossum, chicken, and frog; teleost: Rainbow trout, carp, and loach. The most commonly used codon for each amino acid is shown in parentheses.

The coelacanth O_L loop, unlike other fish which have a polypyrimidine tract, but similar to tetrapods, contains a stretch of thymines that is needed for the initiation of L-strand synthesis (WONG and CLAYTON 1985).

Coding sequences: The coelacanth *12S* and *16S* rRNA genes are 983 and 1665 nucleotides long, respectively (Table 2). The secondary structure of both rRNA genes appears to be reasonably conserved in respect to those of other vertebrates. Most of our rRNA gene sequence showed only minor differences (99.58% similarity) to that previously reported (HEDGES et al. 1993) with the exception of the last 332 bp of the *16S* rRNA gene sequence, which differed by 20%. These 332 bp correspond to the last PCR fragment amplified by HEDGES et al. (1993) and a search in GenBank revealed that the sequence of this PCR fragment had 96.7% similarity to that sequence of the alligator *16S* rRNA gene sequence (HEDGES 1994). This and other reported cases of apparently scrambled sequences during data preparation can be recognized and prevented by individually checking PCR amplified fragments for compatibility in recovering congruent phylogenetic trees (see EDWARDS and ARCTANDER, 1997).

All 22 coelacanth tRNA gene sequences can be folded into a canonical cloverleaf secondary structure with the exception of the tRNA$^{Ser(AGY)}$ (data not shown). Two mismatched base pairs in the stems of these putative cloverleaf secondary structures were detected on average for each tRNA. The coelacanth tRNAs ranged in size from 65 to 76 nucleotides (Table 2) and showed size variability in their DHU and TψC arms when compared with other vertebrate mitochondrial tRNAs.

The coelacanth mitochondrial genome contains 13 protein-coding genes (Figures 1 and 2); in two cases there is a reading-frame overlap on the same strand (*ATPases 8* and *6* share 10 nucleotides; *ND4L* and *ND4* overlap by seven nucleotides). All coelacanth mitochondrial protein-coding genes begin with a ATG start codon except *COI*, which starts with GTG (Table 2). This initiation codon usage is also found in all other fish mitochondrial genomes that have been completely sequenced so far (TZENG et al. 1992; CHANG et al. 1994; LEE and KOCHER 1995; ZARDOYA et al. 1995; JOHANSEN and BAKKE 1996; NOACK et al. 1996; ZARDOYA and MEYER 1996a). Most coelacanth ORFs end with TAA (*ND1*, *ND2, COI, ATPase 8, COIII, ND4L* and *ND5*), one ends with AGA (*ND6*), and the rest have incomplete stop codons, either T (*COII, ND3, ND4*) or TA (*Cyt b*) (Table 2).

The codon usage of the coelacanth is similar to that of lamprey (LEE and KOCHER 1995), bichir (NOACK et al. 1996), carp, trout and loach (TZENG et al. 1992; CHANG et al. 1994; ZARDOYA et al. 1995), lungfish (ZARDOYA and MEYER 1996a), and tetrapods (e.g., ROE et al. 1985; DESJARDINS and MORAIS 1990; ARNASON and GULLBERG 1993; HORAI et al. 1995) (Table 4). A total of 3803 amino acids are encoded by the coelacanth mitochondrial genome. As in other vertebrates, the most abundant amino acid residue encoded by the coelacanth mitochondrial genome is leucine, whereas the
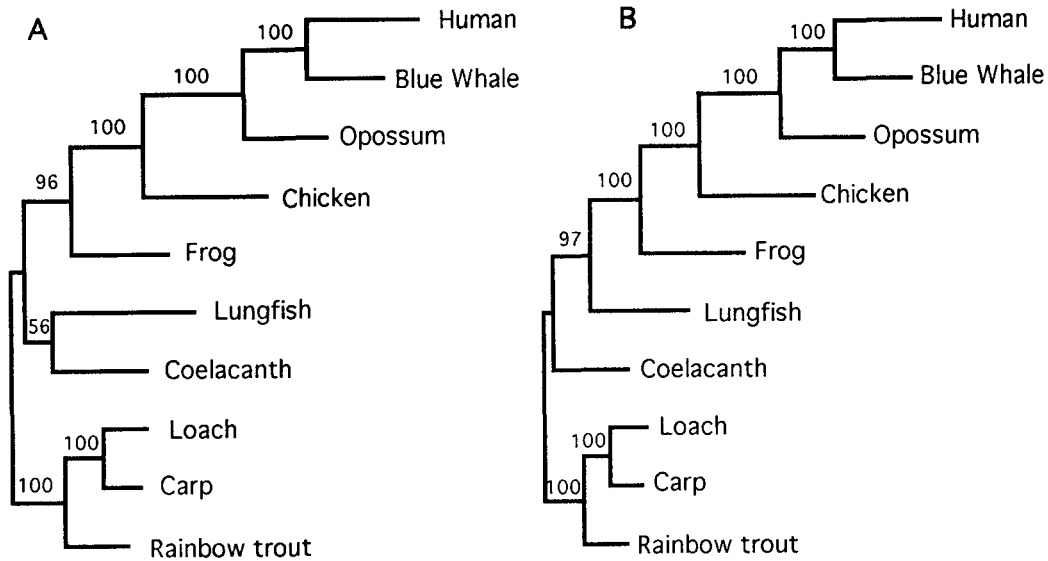
FIGURE 3.—Phylogenetic position of the coelacanth. A data set combining all (protein-coding, rRNA and tRNA) mitochondrial genes was analyzed with maximum parsimony (A), neighbor-joining and maximum likelihood (B) phylogenetic methods. Numbers shown above branches represent bootstrap values from 100 pseudoreplicates for MP (A) and NJ (B). Rainbow trout, carp, and loach were used as outgroup taxa. Third codon positions and transitions in first codon positions of protein coding genes were excluded from the MP phylogenetic analysis. The NJ analysis was performed using Kimura two-parameter distances, an α = 0.53 (gamma distribution), and excluding third codon positions. The ML analysis was performed with a transition: transversion ratio of 2:1, using the F84 model implemented in PHYLIP version 3.55 (FELSENSTEIN 1989), which accounts for different base frequencies, and excluding third codon positions.

rarest is cysteine (Table 4). Adenines and cytosines are preferentially used in third codon positions of the coelacanth mitochondrial protein coding genes, indicating that codon usage is probably influenced by strand-specific base composition mutational bias (FLOOK et al. 1995; JERMIIN et al. 1994; LEE and KOCHER 1995). Alternatively, third codon position usage may be determined by availability of ribonucleotides in the mitochondria (XIA 1996).

Phylogenetic position of the coelacanth: To asses the position of the coelacanth with respect to lungfishes and tetrapods, a combined data set comprising rRNA, tRNA and protein coding nucleotide sequences of the human (HORAI et al. 1995), 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), rainbow trout (ZARDOYA et al. 1995), African lungfish (ZARDOYA and MEYER 1996a), and coelacanth mitochondrial genomes was constructed; it is composed of a total of 16,140 characters. This data set was analyzed with the three most commonly used methods of phylogenetic inference, i.e., maximum parsimony (MP; FITCH 1971), neighbor-joining (NJ; SAITOU and NEI 1987), and maximun likelihood (ML; FELSENSTEIN 1989).

When third codon positions and transitions in first codon positions of the protein-coding gene sequences were excluded from the analysis and teleosts (trout, carp and loach) were included as outgroup taxa, MP recovered one most parsimonious tree (8468 steps, C.I. = 0.65) in which a coelacanth/lungfish clade is identi-

fied as the sister group of tetrapods (Figure 3A). This is the same hypothesis as that which is favored by the largest available nuclear DNA data set, the complete *28S rRNA* gene sequences (ZARDOYA and MEYER 1996b). The node grouping the coelacanth/lungfish clade with tetrapods was supported by a 56% bootstrap value (Figure 3A). However, if the same data set was analyzed with MP and transitions in first positions of the protein-coding genes were included in the analysis or a transversion:transition weighting of 2:1 was assumed for the whole data set, a lungfish/tetrapod clade was favored with low or moderate bootstrap support (51 and 68%, respectively).

For the NJ analysis, to account for the variation of substitution rates among sites, the α shape parameter of the gamma distribution of rate variation was estimated based on the MP tree (Figure 3A) by the method of YANG and KUMAR (1996). A NJ analysis of the same data set, excluding third codon positions of the protein-coding genes, with teleosts as outgroup taxa, using Kimura 2-parameter distances, and an α = 0.53, arrived at a tree in which lungfishes are placed as the sister group of tetrapods (97% bootstrap value for the lungfish+tetrapod node) (Figure 3B). The same topology was recovered when a ML analysis (F84 model, 2:1 transition:transversion ratio; Ln likelihood = −61,846.08) was performed excluding third codon positions of the protein-coding genes and including teleosts as outgroup taxa. In this case, all branch lengths were found to be significantly greater than zero ($P <$ 0.01).

The disagreement between methods of phylogenetic inference was investigated with KISHINO and HASEGAWA (1989) and TEMPLETON (1983) tests. None of the three possible hypotheses, *i.e.,* lungfish as sister group of tetrapods (Ln likelihood = $-61,846.08$, best ML tree; 10,812 steps, best MP tree), coelacanth as sister group of tetrapods [Ln likelihood = $-61,863.43$, ΔLnL = 17.35 $\pm$ 22.77 (SD); 10,832 steps, Δsteps = 20 $\pm$ 15.43], or coelacanth as sister group of lungfish (Ln likelihood = $-61,859.78$, ΔLnL = 13.70 $\pm$ 22.69; 10,813 steps, Δsteps = 1 $\pm$ 16.03), could be statistically rejected based on the mitochondrial DNA data set when third codon positions of mitochondrial protein-coding genes were excluded from the analyses and teleosts were used as outgroup taxa. In a separate publication, we will report in more detail on the conflicting phylogenetic signal that is contained in the tRNA, rRNA, and protein mitochondrial data sets and that may be one of the major causes that explain why none of the three hypotheses can be statistically ruled out (R. ZARDOYA, Y. CAO, M. HASEGAWA and A. MEYER,unpublished data).

Our evolutionary analyses confirmed the phylogenetic position of the coelacanth as a closer relative to tetrapods and other sarcopterygians than to the ray-fin fishes (Actinopterygii). These results indicate that the rate of evolution of the mitochondrial genome is appropriate for resolving relationships even among ancient lineages (at least up to the Devonian). However, it seems to fall short of recovering and strongly supporting the relationships among the extant lineage of lobe-finned fishes that originated within a narrow window in time ~400 mya. The rapid origin of the various lobe-finned lineages and their rapid radiation before the origin of tetrapods continues to make it difficult to resolve their relationships with confidence.

LITERATURE CITED

ADACHI, J., and M. HASEGAWA, 1992 *MOLPHY: Programs for Molecular Phylogenetics I-PROTML: Maximum Likelihood Inference of Protein Phylogeny.* Institute of Statistical Mathematics, Tokyo.

AGASSIZ, L., 1844 *Recherches sur les Poissons Fossiles,* Imprimerie de Petitpierre, Neuchatel.

AHLBERG, P. E., 1991 A re-examination of sarcopterygian interrela-

tionships, with special reference to the porolepiformes. Zool. J. Linn. Soc. **103:** 241–288.

AHLBERG, P. E., J. A. CLACK and E. LUKSEVICS, 1996 Rapid braincase evolution between *Panderichthys* and the earliest tetrapods. Nature **381:** 61–63.

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

BEMIS, W. E., and R. G. NORTHCUTT, 1991 Innervation of the basicranial muscle of *Latimeria chalumnae.* Environ. Biol. Fishes **32:** 147–158.

BETZ, U. A., W. E. MAYER and J. KLEIN, 1994 Major histocompatibility complex class I genes of the coelacanth *Latimeria chalumnae.* Proc. Natl. Acad. Sci. USA **91:** 11065–11069.

BRUTON, M. N., and S. E. COUTOUVIDIS, 1991 An inventory of all known specimens of the coelacanth *Latimeria chalumnae,* with comments on trends in the catches. Environ. Biol. Fishes **32:** 371–390.

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

CHANG, M. M., 1991 Rhipidistians, pp. 1–28 in *Origins of the Higher Groups of Tetrapods. Controversy and Consensus,* edited by H. P. SCHULTZE and L. TRUEB. Cornell University Press, Ithaca, NY.

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.

CLACK, J. A., 1994 Earliest known tetrapod braincase and the evolution of the stapes and fenestra ovalis. Nature **369:** 392–394.

CLOUTIER, R., and P. E. AHLBERG, 1996 Interrelationships of basal sarcopterygians, pp. 445–479 in *Interrelationships of Fishes,* edited by M. L. J. STIASSNY, L. R. PARENTI and G. D. JOHNSON. Academic Press, San Diego.

CLOUTIER, R., and P. L. FOREY, 1991 Diversity of extinct and living actinistian fishes (Sarcopterygii). Environ. Biol. Fishes **32:** 59–74.

COURTENAY-LATIMER, M., 1979 My story of the first coelacanth. Occas. Pap. Calif. Acad. Sci. **134:** 6–10.

CUMMINGS, M. P., S. P. OTTO and J. WAKELEY, 1995 Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. **12:** 814–822.

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, J. N., 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.

EDWARDS, S. V., and P. ARCTANDER, 1997 Congruence and phylogenetic re-analysis of perching bird *cytochrome b* sequences. Mol. Phyl. Evol. **7:** 266–271.

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.

FITCH, W.M., 1971 Toward defining the course of evolution: minimal change for a specific tree topology. Syst. Biol. **20:** 406–416.

FLOOK, P. K., C. H. F. ROWELL and G. GELLINSEN, 1995 The sequence organization and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. **41:** 928–941.

FORAN, D. R., J. E. HIXSON and W. M. BROWN, 1988 Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res. **16:** 5841–5861.

FOREY, P. L., 1987 Relationships of lungfishes. J. Morphol. **1** Suppl.: 75–91.

FOREY, P. L., 1988 Golden jubilee for the coelacanth *Latimeria chalumnae.* Nature **336:** 727–732.

FOREY, P. L., 1991 *Latimeria chalumnae* and its pedigree. Environ. Biol. Fishes **32:** 75–97.

FRICKE, H., 1992 Coelacanth tissue bank. Nature **357:** 105.

FRICKE, H., K. HISSMANN, J. SCHAUER and R. PLANTE, 1995 Yet more danger for coelacanths. Nature **374**: 314.

FRITZSCH, B., 1987 Inner ear of the coelacanth fish *Latimeria* has tetrapod affinities. Nature **327**: 153–154.

GELFAND, D. H. and WHITE, T.J. 1990 Thermostable DNA polymerases, pp. 129–141 in *PCR Protocols: A Guide to Methods and Applications*, edited by M. A. INNIS, D. H. GELFAND, J. J. SNISNKY and T. J. WHITE. Academic Press, San Diego.

GORR, T., T. KLEINSCHMIDT and H. FRICKE, 1991 Close tetrapod relationships of the coelacanth *Latimeria* indicated by haemoglobin sequences. Nature **351**: 394–397.

HEDGES, S. B., 1994 Molecular evidence for the origin of birds. Proc. Natl. Acad. Sci. USA **91**: 2621–2624.

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

HILLIS, D. M., M. T. DIXON and L. K. AMMERMAN, 1991 The relationships of the coelacanth *Latimeria chalumnae*: evidence from sequences of vertebrate 28S ribosomal RNA genes. Environ. Biol. Fishes **32**: 119–130.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE and N. TAKAHATA, 1995 Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92**:532–536.

HUXLEY, T. H., 1861 Preliminary essay upon the systematic arrangement of the fishes of the Devonian epoch. Figures and descriptions illustrative of British organic remains. Mem. Geol. Surv. U.K. Dec. **10**: 1–40.

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.

JERMIIN, L. S., D. GRAUR, R. M. LOWE and R. H. CROZIER, 1994 Analysis of directional mutation pressure and nucleotide content in mitochondrial *cytochrome b* genes. J. Mol. Evol. **39**: 160–173.

JERMIIN, L., D. GRAUR and R. H. CROZIER, 1995 Evidence from analyses of intergenic regions for strand-specific directional mutation pressure in metazoan mtDNA. Mol. Biol. Evol. **12**: 558–563.

JOHANSEN, S., and I. BAKKE, 1996 The complete mitochondrial DNA sequence of Atlantic cod (*Gadus morhua*): relevance to taxonomic studies among codfishes. Mol. Mar. Biol. Biotech. **5**: 203–214.

KISHINO, H., and M. HASEGAWA, 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.

KOCHER, T. D., W. K. THOMAS, A. MEYER, S. V. EDWARDS, S. PÄÄBO *et al.*, 1989 Dynamics of mitochondrial DNA evolution in animals. Proc. Natl. Acad. Sci. USA **86**: 6196–6200.

KOLB, E., J. I. HARRIS and J. BRIDGEN, 1974 Triose phosphate isomerase EC-5.3.1.1 from the coelacanth: an approach to the rapid determination of an amino-acid sequence with small amounts of material. Biochem. J. **137**: 185–197.

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.

LONG, J. A., 1995 *The Rise of Fishes: 500 Million Years of Evolution.* The John Hopkins University Press, Baltimore.

MADDISON, W. P., and D. R. MADDISON, 1992 *MacClade: Analysis of Phylogeny and Character Evolution.* Sinauer Associates, Sunderland, MA.

MAISEY, J. G., 1996 *Discovering Fossil Fishes.* Holt, New York.

MANGUM, C. P., 1991 Urea and chloride sensitivities of coelacanth hemoglobin. Environ. Biol. Fishes **32**: 219–222.

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

MILLOT, J., 1954 New facts about coelacanths. Nature **174**: 426–427.

MILLOT, J., 1955 The coelacanth. Sci. Am. **193**: 34–39.

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

NOACK, K., R. ZARDOYA and A. MEYER, 1996 The complete mitochondrial DNA sequence of the bichir (*Polypterus ornatipinnis*), a basal ray-finned fish: ancient establishment of the consensus vertebrate gene order. Genetics **144**: 1165–1180.

NORTHCUTT, R. G., 1987 Lungfish neural characters and their bearing on sarcopterygian phylogeny. J. Morphol. **1** (Suppl.): 277–297.

PALUMBI, S.R., A. MARTIN, S. ROMANO, W. O. MCMILLAN, L. STICE *et al.*, 1991 *The Simple Fool's Guide to PCR.* Department of Zoology, University of Hawaii, Honolulu.

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

PECHERE, J. F., H. ROCHAT and C. FERRAZ, 1978 Parvalbumins from coelacanth muscle. II Amino acid sequence of the two less acidic components. Biochim. Biophys. Acta **536**: 269–274.

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. Natl. Mus. Nat. Hist. **167**: 159–276.

RUSSO, C. A. M., N. TAKEZAKI and M. NEI, 1996 Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. **13**: 525–536.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406–425.

SCHLIEWEN, U., H. FRICKE, M. SCHARTL, J. T. EPPLEN and S. PÄÄBO, 1993 Which home for the coelacanth? Nature **363**: 406.

SCHULTZE, H. P., 1987 Dipnoans as sarcopterygians. J. Morphol. **1** (Suppl.): 39–74.

SCHULTZE, H. P., 1994 Comparison of hypotheses on the relationships of sarcopterygians. Syst. Biol. **43**: 155–173.

SCHULTZE, H. P., and R. CLOUTIER, 1991 Computed tomography and magnetic resonance imaging studies of *Latimeria chalumnae.* Environ. Biol. Fishes **32**: 159–181.

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

SMITH, J. L. B., 1939 A living fish of mesozoic type. Nature **143**: 455–456.

SMITH, J. L. B., 1956 *Old Fourlegs: The Story of the Coelacanth.* Longmans, 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.

STENSIÖ, E. A., 1921 *Triassic Fishes from Spitzbergen.* Holzhausen, Vienna.

STOCK, D. W., K. D. MOBERG, L. R. MAXSON and G. S. WHITT, 1991 A phylogenetic analysis of the 18S ribosomal RNA sequence of coelacanth *Latimeria chalumnae.* Environ. Biol. Fishes **32**: 99–117.

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

SWOFFORD, D. L., 1997 *PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods)*, version 4.0. Sinauer Associates, Sunderland, MA.

TAMAI, Y., H. KOJIMA and K. TAKAYAMA-ABE, 1994 Lipids and myelin proteins in the brains of coelacanth (*Latimeria chalumnae*), lungfish (*Lepidosiren paradoxa* and *Protopterus aetiopicus*), bichir (*Polypterus senegalus*), and sturgeon (*Acipenser ruthenus*) (Osteichthyes): phylogenetic implications. Can. J. Fish. Aquat. Sci. **51**: 1265–1272.

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

THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673–4680.

THOMSON, K. S., 1966 Intercranial mobility in the coelacanth. Science **153**: 999–1000.

TOWNER, P., 1991   Purification of DNA, pp. 47–68 in *Essential Molecular Biology. A Practical Approach*, edited by T. A. BROWN, Oxford University Press, Oxford.

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.

VOROBYEVA, E., and H. P. SCHULTZE, 1991   Description and systematics of panderichthyid fishes with comments on their relationship to tetrapods, pp. 68–109 in *Origins of the Major Groups of Tetrapods: Controversies and Consensus*, edited by H. P. SCHULTZE, and L. TRUEB, Cornell Univ. Press, Ithaca, NY.

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.

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.

WOODWARD, A. S., 1891   *Catalogue of Fossil Fishes in the British Museum (Natural History)*. British Museum (Natural History), London.

XIA, X., 1996   Maximizing transcription efficiency causes codon usage bias. Genetics **144:** 1309–1320.

XU, X. and U. ARNASON, 1996   A complete sequence of the mitochondrial genome of the Western lowland gorilla. Mol. Biol. Evol. **13:** 691–698.

YANG, Z., and S. KUMAR, 1996   Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. Mol. Biol. Evol. **13:** 650–659.

YOKOBORI, A. I., M. HASEGAWA, 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., and A. MEYER, 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., and A. MEYER, 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., and A. MEYER, 1996c   Phylogenetic performance of mitochondrial protein coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. **13:** 933–942.

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

Communicating editor: N. TAKAHATA