The Complete Mitochondrial DNA Sequence of the Bichir (*Polypterus ornatipinnis*), a Basal Ray-Finned Fish: Ancient Establishment of the Consensus Vertebrate Gene Order

Katharina Noack, 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 evolutionary position of bichirs is disputed, and they have been variously aligned with ray-finned fish (Actinopterygii) or lobe-finned fish (Sarcopterygii), which also include tetrapods. Alternatively, they have been placed into their own group, the Brachiopterygii. The phylogenetic position of bichirs as possibly the most primitive living bony fish (Osteichthyes) made knowledge about their mitochondrial genome of considerable evolutionary interest. We determined the complete nucleotide sequence (16,624 bp) of the mitochondrial genome of a bichir, *Polypterus ornatipinnis*. Its genome contains 13 protein-coding genes, 22 tRNAs, two rRNAs and one major noncoding region. The genome's structure and organization show that this is the most basal vertebrate that conforms to the consensus vertebrate mtDNA gene order. Bichir mitochondrial protein-coding and ribosomal RNA genes have greater sequence similarity to ray-finned fish than to either lamprey or lungfish. Phylogenetic analyses suggest the bichir's placement as the most basal living member of the ray-finned fish and rule out its classification as a lobe-finned fish. Hence, its lobe-fins are probably not a shared-derived trait with those of lobe-finned fish (Sarcopterygii).

BONY fish (Osteichthyes) are typically divided into two major groups, the actinopterygians (ray-finned fish) and the sarcopterygians (lobe-finned fish). With >25,000 known species, the ray-finned fish comprise by far the largest class of vertebrates. Ray-finned fish date back to the early Devonian (CARROLL 1988), and their diversity has since then increased steadily both in terms of morphological diversity and in terms of numbers of species. By contrast, the once highly successful and diverse lobe-finned fish, now only include two extant fish groups, the lungfish and the coelacanth. During the last 150 yr, both of these latter groups have been variously thought to be the closest living relatives of land vertebrates (*e.g.*, ZARDOYA and MEYER 1996a; reviewed in MEYER 1995).

While the classification of bony fish into either the ray-finned or the lobe-finned fish is without problems, the phylogenetic placement of species of the order Polypteriformes into one of these two categories (*e.g.*, PATTERSON 1982; LAUDER and LIEM 1983, also referred to as Cladistia) has been historically difficult, and it still remains somewhat uncertain (Table 1). Unlike any other group of living fish (reviewed in PATTERSON 1982; LAUDER and LIEM 1983), polypteriform fish (bichirs, Polypterus, and reedfish, Calamoichthys) have been placed into a number of widely differing taxonomic groups during the last century.

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

Figure 1 summarizes some of the various phylogenetic hypotheses involving bichirs. One of the earliest hypotheses of polypterid ancestry held that they were sarcopterygians, and bichirs were placed into the suborder Crossopterygii along with fossil rhipidistians, coelacanths and lungfish (Figure 1a) (HUXLEY 1861; COPE 1871; reviewed in PATTERSON 1980, 1982). Goodrich (1928), for the first time, considered placing polypterids within the actinopterygians as a taxonomically distinct lower division. Since then, the alternative two main hypotheses of polypterine relationships have been that polypterids are either classified as a distinct subclass of Osteichthyes, the Brachiopterygii (Figure 1b), or as basal actinopterygians (Figure 1c). The main problems of assessing polypterine relationships are that the fossil record for these fish only extends back to the Eocene (GREENWOOD 1974; CARROLL 1988) and that they show a large number of primitive ancestral as well as uniquely derived characteristics that make the determination of their phylogenetic position, and hence their classification, difficult. Bichirs have a row of dorsal finlets that extend down their back, primitive rhombic ganoid scales, numerous spiracular plates, and lobed pectoral fins. Their pectoral fins have a fleshy external appearance similar to those of lungfish and the coelacanth. Polypterine fish also have a well-vascularized lung that arises from the ventral foregut and a pair of external gills in juveniles, a rare feature, not seen in any other fish, except in some species of lungfish.

Many authors still argue that Polypteriformes are

TABLE 1

Systematic position of bichirs

Superclass: Gnathostomata
  Class Chondrichthyes
    Subclass: Holocephali (Chimaeras)
    Subclass: Elasmobranchii (sharks, rays)
  Class Sarcopterygii
    Subclass: Coelacanthimorpha (Latimeria)
    Subclass: Dipnoi (lungfishes)
    Subclass: Tetrapoda (amphibians, reptiles, birds,
      mammals)
  Class Actinopterygii
    Subclass: Chondrostei
      Order: Polypteriformes (bichirs, Polypterus,
        reedfishes, Calamoichthys)
      Order: Acipenseriformes (sturgeons, Acipenser;
        paddlefishes, Polyodont)
    Subclass: Neopterygii
      Order: Semionotiformes (gars, Lepisosteus)
      Order: Amiiformes (bowfins, Amia)
        Division: Teleostei

Modified from NELSON (1994).

morphologically too distinct from actinopterygians to be classified as such (JESSEN 1973; NELSON 1973; JARVIK 1980; BJERRING 1985). This view, however, does not clarify what other group of fish polypterids are most closely related to. Most recent authors have, based on morphological characters such as scale structure and neurocranial ossification, placed polypterids with actinopterygians (Figure 1c) (GARDINER 1973; SCHAEFFER 1973; PATTERSON 1980, 1982; LAUDER and LIEM 1983; GARDINER and SCHAEFFER 1989). However, the placement of Polypteriformes within the Actinopterygii is also uncertain. Often polypterids are placed within the Chondrostei, a group to which most authors also add the sturgeons and their relatives (Table 1, e.g., GARDINER 1967; ANDREWS 1967; CARROLL 1988; NELSON 1994). Whereas other researchers classify polypterids as a taxon of equal rank and basal to the Chondrostei (reviewed in PATTERSON 1982; LAUDER and LIEM 1983).

Among other reasons, knowledge of the phylogenetic position of the bichir is of importance for the reconstruction of the development and evolution of verte-

A

HUXLEY 1861
COPE 1871

B

JESSEN 1973
NELSON 1973
JARVIK 1980
BJERRING 1985
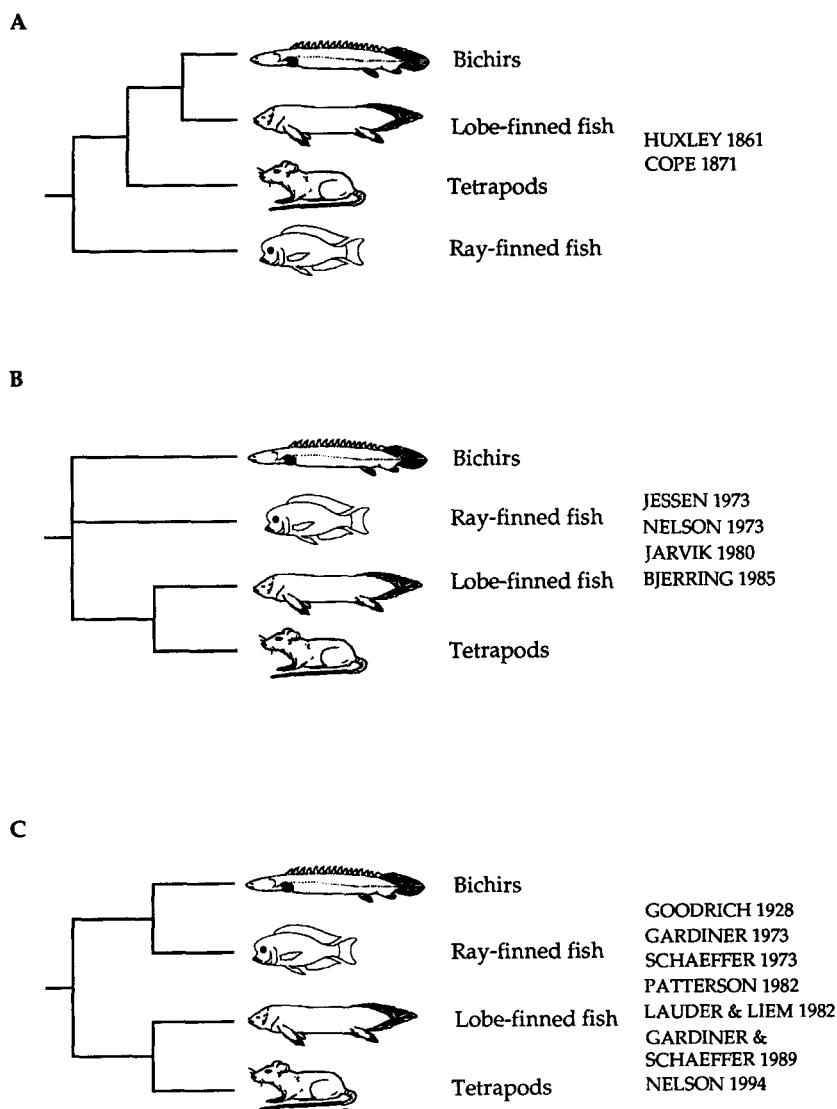
FIGURE 1.—Alternative hypotheses of relationships between bichir, ray-finned fish and lobe-finned fish. (A) Polypteriformes are lobe-finned fish. (B) Polypteriformes are a third subclass of Osteichthyes with an unresolved position in respect with ray-finned and lobe-finned fish. (C) Polypteriformes are primitive ray-finned fish.

C

GOODRICH 1928
GARDINER 1973
SCHAEFFER 1973
PATTERSON 1982
LAUDER & LIEM 1982
GARDINER &
SCHAEFFER 1989
NELSON 1994

brate traits such as paired limbs from fins (e.g., SHUBIN and ALBERCH 1986; SORDINO and DUBOULE 1996). However, so far, only two molecular phylogenetic studies have incorporated nucleotide sequence data from bichirs. NORMARK et al. (1991) used partial DNA sequences of the mitochondrial cyt b, COI and COII genes to infer phylogenetic relationships of Neopterygians, using the bichir as outgroup taxon. Lê et al. (1993) used portions of the large nuclear ribosomal gene in an attempt to infer the phylogeny among major fish lineages. Their study supported the view that polypterids are a primitive ray-finned fish lineage, albeit with low confidence.

Mitochondrial genes are increasingly used to infer phylogenetic relationships both among closely related species as well as distantly related ones (KUMAZAWA and NISHIDA 1993; CAO et al. 1994; RUSSO et al. 1996; ZARDOYA and MEYER 1996c). ZARDOYA and MEYER (1996a) show that combined sets of all mitochondrial protein coding and tRNA genes can with strong support resolve the relationships among major groups of jawed vertebrates. Until now the complete mitochondrial genome sequences of 22 vertebrate species have been reported; 15 are from mammals, five from fish, only one from an amphibian and one from a bird. The structure and organization of vertebrate mitochondrial genomes is remarkably conserved. Minor rearrangements, apparently caused by the translocation of tRNA genes, have been described for chicken (DESJARDINS and MORAIS 1990) and for the opposum (JANKE et al. 1994). It is not yet clear when the establishment of the vertebrate consensus gene order occurred during their evolution. Lampreys, one of the earliest vertebrates, have a peculiar gene order (LEE and KOCHER 1995), but without the study of more chordates, it is impossible to determine if this gene order represents an ancestral or a uniquely derived condition. We determined the complete nucleotide sequence and gene order of the bichir mitochondrial genome with the aim to investigate the evolution of the vertebrate mitochondrial genome and to clarify the bichir's relationships to lobe-finned and ray-finned fish. Furthermore, we assessed the performance of the bichir as an outgroup taxon for inferring relationships among major groups of jawed vertebrates.

MATERIALS AND METHODS

Mitochondrial DNA was purified from fresh liver and kidney tissue of commercially obtained bichir (Polypterus ornatipinnis) as previously described (ZARDOYA et al. 1995). After homogenization, intact nuclei and cellular debris were removed by low-speed centrifugation (1000 × g). Mitochondria were pelleted by centrifugation at 10,000 × g for 20 min and subjected to a standard alkaline lysis procedure. The isolated mtDNA was cleaved with EcoRI and HindIII restriction enzymes. Three EcoRI fragments of 4.0, 3.9 and 3.0 kb and four HindIII fragments of 3.1, 1.0, 0.6, and 0.3 kb (See Figure 2 for positions of restriction sites) were cloned into pUC18 covering the entire bichir mtDNA molecule except for a stretch of 3.1 kb spanning the COI and COII region. This

3.1-kb fragment was amplified using specific oligonucleotide primers and then cloned into pGEM-T (Promega).

Plasmid DNA was extracted from each clone using a Wizard miniprep kit (Promega). After ethanol precipitation, purified 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 (373A Stretch, Applied Biosystems Inc.). Sequences were obtained using both M13 universal sequencing primers and 37 specifically designed oligonucleotide primers. The sequences obtained from each clone were ~350 bp in length and each sequence overlapped the next by ~100 bp. In no case were differences in sequence observed among the overlapping regions.

Sequence data were analyzed by use of the GCG program package (DEVEREUX et al. 1984) and alignments were performed using CLUSTAL W (THOMPSON et al. 1994). The data were subjected to maximum parsimony (MP) analyses using PAUP Version 3.1.1 (SWOFFORD 1993). MP analyses were performed using heuristic searches (TBR branch swapping, MULPARS option in effect, simple stepwise addition of taxa). Maximum likelihood (ML) analyses and neighbor-joining (NJ) analyses were performed using PHYLIP Version 3.5 (FELSENSTEIN 1989). For NJ analyses, distance matrices were calculated based on Kimura corrected distances (KIMURA 1980). The Kishino-Hasegawa test to assess the statistical confidence of maximum likelihood trees was performed using MOLPHY Version 2.2 (ADACHI and HASEGAWA 1992). Confidence levels for MP and NJ analyses were assessed by bootstrap analyses based on 1000 replications (FELSENSTEIN 1985) using PHYLIP Version 3.5 (FELSENSTEIN 1989).

The complete mtDNA sequence of the bichir has been deposited in GenBank under the accession No. U62532.

RESULTS AND DISCUSSION

Genome organization: The total length of the bichir mitochondrial genome is 16,624 bp, it is similar to that of the mitochondrial genomes of other bony fish (Table 2). The overall base composition of the L-strand is A = 32.2%, T = 28.1%, C = 25.6% and G = 14.1%. As found in other vertebrates, the bichir's mitochondrial genome contains 13 protein coding genes, 22 transfer RNA genes, and two ribosomal RNA genes. Interestingly, the orientation and relative position of all genes and the major noncoding region is identical to the vertebrate consensus mitochondrial gene order (Figures 2 and 3, Table 3). The presence of this gene order in the bichir documents the early establishment of the particular vertebrate organization of the mitochondrial genome. All genes except the gene encoding for ND6 and eight transfer RNAs are encoded on the H-strand, and the genes encoded on both strands of the bichir mtDNA are of similar length as those in other vertebrates (Table 3). Peptide encoding genes were identified by comparison with the rainbow trout mtDNA (ZARDOYA et al. 1995) and by the presence of initiation and stop codons. Transfer RNA genes were identified by sequence similarity to homologous vertebrate tRNAs, their specific anticodons, and their ability to fold into putative cloverleaf structures.

As seen in other vertebrates there is an overlap between the genes encoding for ND4 and ND4L, ATPase

TABLE 2

Lengths of fish mitochondrial genes

	Species					
Gene	Lungfish	Trout	Carp	Loach	Bichir	Lamprey
Control region	1184	1003	927	896	1068	491
12S rRNA	937	944	951	937	950	900
16S rRNA	1591	1680	1681	1680	1655	1621
Cytb	1144	1141	1141	1144	1141	1191
ND1	966	972	975	975	958	966
ND2	1028	1050	1047	1047	1036	1044
ND3	346	349	351	351	346	351
ND4	1384	1381	1383	1383	1378	1377
ND4L	297	297	297	297	297	291
ND5	1836	1839	1824	1837	1842	1797
ND6	513	522	519	522	504	519
COI	1548	1551	1551	1551	1557	1554
COII	691	691	691	691	688	690
COIII	784	784	786	768	784	786
APT6	682	670	684	684	682	714
ATP8	168	168	165	168	168	168
Total	16646	16642	16575	16558	16624	16201

Values expressed as base pairs.

6 and ATPase 8, and ND5 and ND6. The bichir mtDNA also has an overlap of nine nucleotides involving *COI* and *tRNA$^{Ser(UCN)}$*, which are encoded on opposite DNA strands. This overlap is also found in the lamprey (LEE and KOCHER 1995) but not in any other fish (TZENG *et al.* 1992; CHANG *et al.* 1994; ZARDOYA *et al.* 1995; ZARDOYA and MEYER 1996a) where these two genes either directly abutt or are separated by a few nucleotides. These data show that the birchir is the most basal vertebrate known so far with the consensus vertebrate mito-

chondrial gene order. Once a complete chondrichthyan mitochondrial genome has been sequenced, the exact phylogenetic timing of the evolution of this gene order in vertebrates will have been established.

Protein coding genes: The bichir mitochondrial genome encodes for 13 proteins. Pairwise comparisons of the protein coding genes of the bichir with their homologues of rainbow trout, loach, carp, lungfish and lamprey showed that bichir proteins consistently have greater sequence similarity to those from ray-finned fish

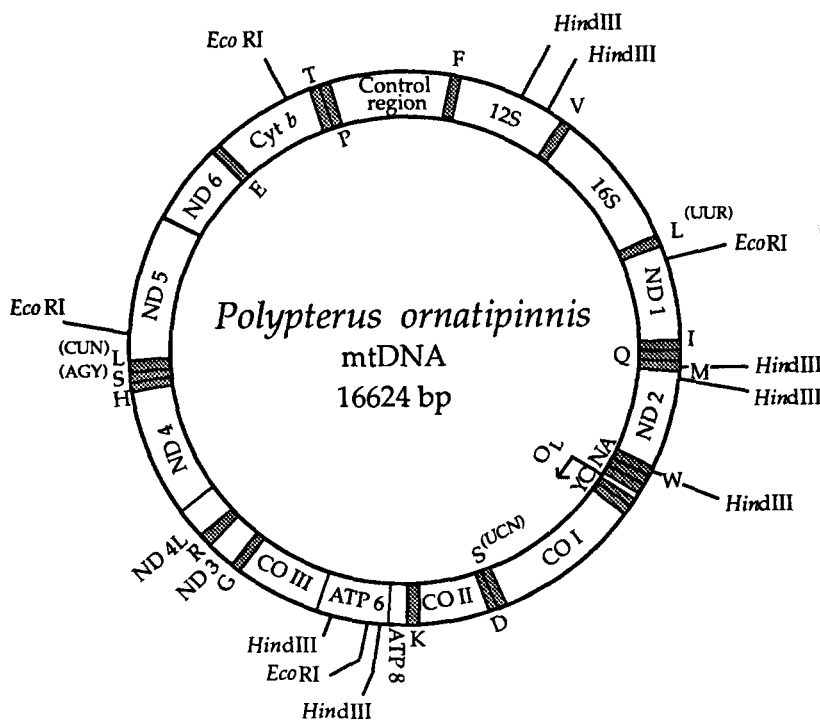FIGURE 2.—Restriction map and gene organization of the *Polypterus ornatipinnis* mitochondrial genome. All protein coding genes are encoded by the H-strand with the exception of *ND6*, which is coded by the L-strand. Each tRNA gene is identified by the single letter amino acid code and depicted according to the coding strand. Only the *Eco*RI and *Hind*III restriction sites used for cloning are shown.

tRNA-Phe→   ___   12S rRNA→

```
  1  GCTAGTGTAGCTTATCCTCAAAGCATAACACTGAAAATGTTAAGACGGACCCTAGAAAAGTCCCACAAGCACAAAGATTTGGTCCTAGTCTTACTATCAT   100

101  TTCTCGACTAAACTTATACATGCAAGTATCCGCGCTCCGGTGAAAATGCCCTCAATCTTCCTAGTAGAGGATAAGGAGCTGGCATCAGGCTCGTGTCCAC   200

201  GGCCCCAAAACGCCTTGCTTTGCCACACCCCCACGGGATTTCAGCAGTAGTAAACATTAGGCAATCAGCGAAAGCTAGACCTAGTTATAGTTAATTAGAG   300

301  CCGGTAAAACTCGTGCCAGCCACCGCGGTTATACGAGGGGCTCCAAATGATGGTTATACGGCGTATGGCGTGGTTAACAATTACAATAAAAACTGAGAAC   400

401  AAAATACCTTTAAGCTGTCATACGCTAATAAGTCCATGAAAATCAACAACGAAAGTTATCTCATTATATGTGAACCCACGAAAACCAAGGCACAAACTGG   500

501  GATTAGATACCCCACTATGCTTGGTCTTAAACTAAGGCGGCACTTATACTAAGCCGCTCGCCAGGTTACTACGAGCGCAAGCTTAAAACCCAAAGGACTT   600

601  GGCGGTGCTTCAGAACCACCTAGAGGAGCCTGTTCTGTAATCGATAATCCTCGTTCAACCTCACCGCATCTTGCATTCCAGCCTATATACCGCCGTCGCC   700

701  AGTCTACCTTTTGAGAGACCATAAGTAGGCCCAATAAGTAAAACTTAACACGTCAGGTCAAGGTGTAGCTTATGATGCGGGAAGAAATGGGCTACATTTT   800

801  CTAGATTAGAATACACGAAAGACACTATGAAACCTGTGTCTGAAGGAGGATTTAGCAGTAAATGGGGAGTAGAGTGCCCACTGAAGCTTGGCGCTGAAGC   900

901  GCGCACACACCGCCCGTCACTCTCCTCCAAACATTTTATAATTTTTAATAAAACACTTCCGACAATAAGAGGAGGCAAGTCGTAACATGGTAAGTGTACC  1000
```

                       tRNA-Val→   ___   16S rRNA→

```
1001  GGAAGGTGCACTTGGAATAACCAAAATGTAGCTTAATAGAAAAGTATCTCCCTTACACTGAGATGAAGTCTGCGCAATACAGACCATTTTGAGCATTATA  1100

1101  GCTAGCCTGACCACACACTAAACCAAACACTATTAAAATAAATATACTTATAATAAAAACAAAACATTTGTTAACTTCAGTATAGGCGATAGAAAGAGAA  1200

1201  CAAAAGAGCTATAGCAACAGTACCGCAAGGGAAAGCTGAAAGAGAAATGAAACAAATCGTTAAAGCACGACATAGCAGAGATTAAATCTCGTACCTTTTG  1300

1301  CATCATGATCTAGTAAGTAAGCCCAAGCAAAATGATTTATAGTTTGACCCCCCGAAACTAGACGAGCTACTTCGAGGCAGTTGAAAGGACCACCCGTCT  1400

1401  CTGTGGAAAAAGAGTGGGAAGACTTCCAAGTAGAGGTGACAAGCCTAACGAGCCTAGTGATAGCTGGTTACTTGAGAAATGGATAAAAGTCCAGCCTCAA  1500

1501  AATTTCTAAAAAATATACAAATTCCGTTTAAAAAATTTTAAGAACATTTGAGAGTTATTCACAGGAGGTACAGCTCCTATGAACTGGGAAACAACCCAAT  1600

1601  AAGGAGGAAAAAGATCATAATTTACAAGGACAAAATCCAAGTGGGCCTGAAAGCAGCCACCTTTAAAGAAAGCGTTATAGCTTAAATAATATATTATTCC  1700

1701  GTATATCCGGATAAAATCTCTGAATCCCCTACAAATATCAAGTTATTCTATTTGAGTAGAAGAAATTATGCTAGAACTAGTAATAAGAAAAATGATTTTC  1800

1801  TCCTAGCACAAGTGTAAGTTAGAACGGACAAACCACTAACAATTAGACGAACCCAACAGAGGGCCAAATAACGCCTATAATAAAAACAAGAAAACCCTAT  1900

1901  TAATCTTATCGTAAATCTTACACAAGAGTGCCTAAAGGAAAGACTAAAAGAGAAAAAAGGAACTCGGCAAATCCGAGCCCCGCCTGTTTACCAAAAACAT  2000

2001  CGCCTTCAGCTTTTCATGTATTGAAGGTCCTGCCTGCCCAGTGACATGAGTTTAACGGCCGCGGTATCCTGACCGTGCAAAGGTAGCGTAATCACTTGTT  2100

2101  CTTTAAATGAGGACTGGTATGAATGGCCCCACGAGGGCTCAACTGTCTCTTTTTCTCGGTCAATTAAACTGATCTCCCCTGTGCAGAAGCGGGCATAAA  2200

2201  GACATAAGACGAGAAGACCCTGTGGAGCTTTAGACTAAATCCAAACACTCCTCACTATATTTTACCGTATAGACAAACACAGCGTTATGGCCATAAAGTC  2300

2301  TTAGGTTGGGGCGACCACTGAGAACAAATAATCCTCAGCGATGATTGAAGCACAGCTTTATAAACTAAGAATGACAATTCAAAGCATCAGGACACCTGAC  2400

2401  ATTAGGATCCAGACTAATCTGATCAACGAACCAAGTTACCCCAGGGATAACAGCGCAATCTTTTCCAAGAGCCCAAATCGACGAAAAGGTTTACGACCTC  2500

2501  GATGTTGGATCAGGACATCCTAATGGTGCATCCGCTATTAAGGGTTCGTTTGTTCAGCGATTAAAGTCCTACGTGATCTGAGTTCAGACCGGAGCAATCC  2600

2601  AGGTCAGTTTCTATCTATGTAGTCATTTATCCTAGTACGAAAGGATTGGATAAATGGGGCCTATATTAAAGACACGCCCCCTTTTAACCTACTGAAGCCA  2700
```

                                          tRNA-Leu (UUR) →   ___

```
2701  AATCAAGTAGATAATAAAAAACATACCCACGCCCTAGAACAGGGTTAGTTAGAGTGGCAGAGCCTGGTAATTGCATGAGACCTAAACCCTCAATCCCAGG  2800
```

             NADH1→

```
                  M   T   L   I   T   L   I   I   N   P   L   M   Y   I   I   P   I   L   L   A   M   A   F   L   T   L
2801  GGTTCAACCCCCCTCTTTAACTATGACCTTAATCACATTAATTATTAACCCACTCATATATATTATCCCCATTCTGCTAGCAATAGCCTTCCTCACTCTA  2900
         V   E   R   K   M   L   G   Y   M   Q   L   R   K   G   P   N   I   V   G   P   Y   G   L   L   Q   P   I   A   D   G   V   K   L   F
2901  GTAGAACGAAAAATATTAGGCTACATGCAACTCCGAAAAGGACCTAACATCGTAGGCCCATACGGTCTCCTCCAACCCATCGCAGACGGTGTAAAACTAT  3000
         I   K   E   P   V   K   P   S   T   S   S   P   T   L   L   T   P   T   L   A   L   T   L   A   L   I   L   W   I   P   L
3001  TTATTAAAGAACCAGTAAAACCATCAACCTCTTCCCCAACACTTTTTCTACTCACCCCGACACTTGCCCTCACCTTAGCCCTCATTTTATGAATTCCTCT  3100
         P   M   P   L   A   L   T   D   L   N   L   T   I   L   F   I   L   A   V   S   S   L   S   V   Y   S   I   L   G   S   G   W   A
3101  TCCTATACCACTAGCCCTCACAGACCTAAATTTAACTATTCTTTTCATTTTAGCCGTATCAAGCCTATCAGTCTATTCAATCTTAGGTTCAGGTTGAGCT  3200
         S   N   S   K   Y   A   L   I   G   A   L   R   A   V   A   Q   T   I   S   Y   E   V   T   L   G   L   I   I   I   S   L   I   M   F
3201  TCAAATTCCAAATATGCCCTAATTGGAGCATTACGAGCAGTAGCTCAGACAATTTCATATGAAGTCACCTTAGGCCTCATCATCATCTCCTTAATTATAT  3300
            T   G   G   F   T   L   T   T   F   N   T   A   Q   E   A   V   W   L   I   L   P   A   W   P   L   A   A   M   W   F   I   S   T
3301  TTACAGGGGGATTCACATTAACCACCTTCAATACCGCACAAGAAGCCGTATGACTTATCCTACCAGCCTGACCACTAGCCGCAATATGATTCATTTCAAC  3400
         L   A   E   T   N   R   A   P   F   D   L   T   E   G   E   S   E   L   V   S   G   F   N   V   E   Y   A   G   G   P   F   A   L
3401  CCTAGCCGAAACAAACCGAGCACCATTTGATCTTACAGAAGGAGAGTCCGAACTAGTATCTGGATTTAATGTTGAATACGCAGGCGGACCATTCGCCCTC  3500
         F   F   L   A   E   Y   T   N   I   L   L   M   N   A   L   S   T   I   L   F   L   G   P   S   F   N   T   L   T   I   N   L   N   W
3501  TTCTTCCTGGCGGAATACACTAACATCTTACTAATAAATGCTCTCTCTACAATTCTATTTCTCGGCCCATCATTCAACACATTAACTATTAACCTTAACT  3600
```
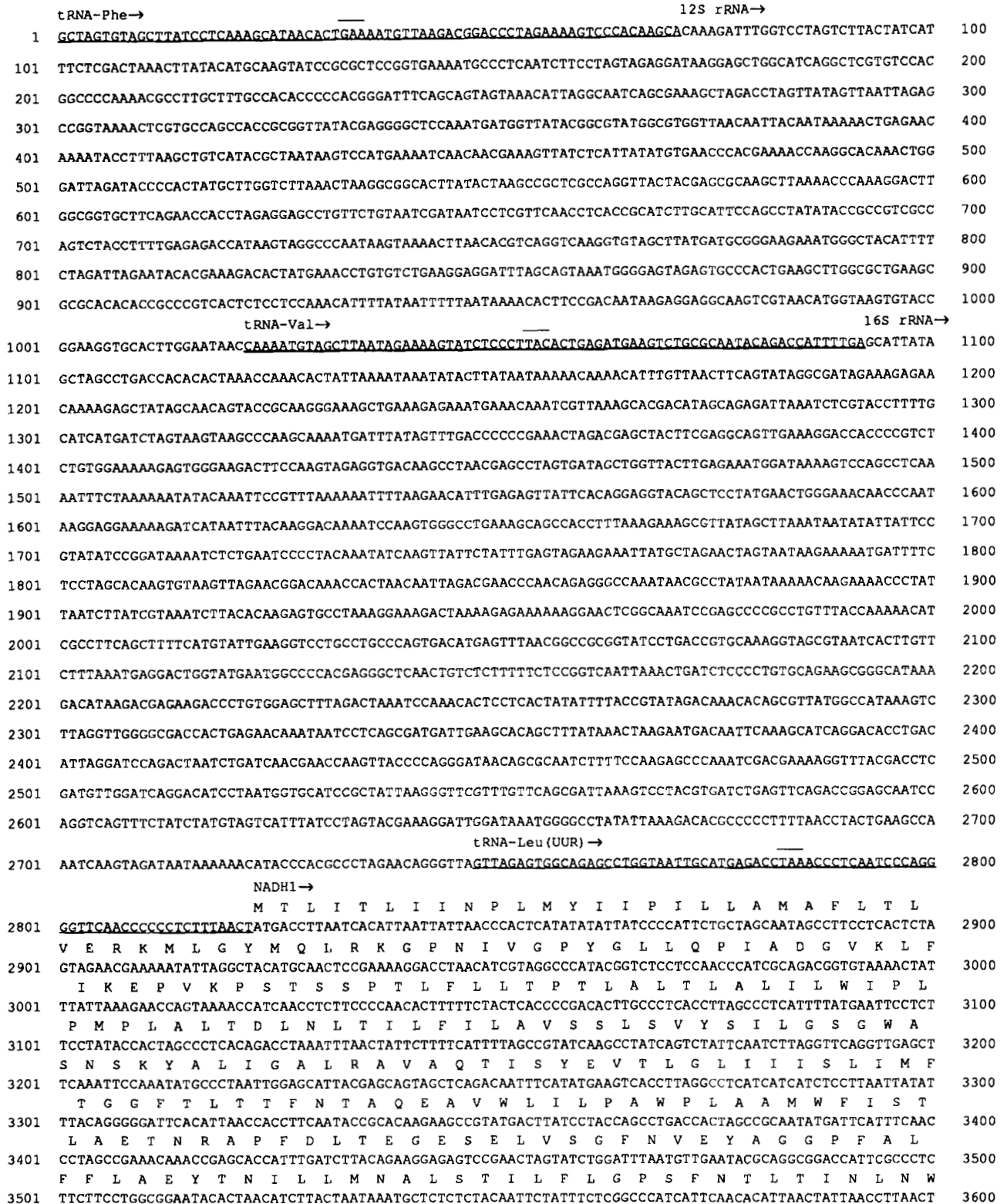
FIGURE 3.—Complete nucleotide sequence of the L-strand of the bichir mitochondrial DNA molecule. Position 1 refers to the first nucleotide of *tRNA*^{*Phe*}. Direction of transcription for each gene is denoted by arrows. The deduced amino acid sequence for each gene product is shown above the nucleotide 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.

```
           A  I  K  T  M  I  L  A  S  M  F  L  W  V  R  A  S  Y  P  R  F  R  Y  D  Q  L  M  H  L  V  W  K  N
3601  GAGCCATCAAAACAATAATCTTAGCATCCATATTCCTCTGAGTTCGCGCCTCCTACCCTCGATTCCGTTATGATCAACTCATACACTTAGTATGAAAAAA  3700
           F  L  P  L  T  L  A  L  I  T  W  H  I  S  L  P  I  S  M  A  G  S  P  P  Q  L      tRNA-Ile→
3701  TTTCCTACCCCTCACTCTTGCTTTAATTACCTGACATATTTCACTACCAATCTCCATGGCAGGTTCACCACCGCAACTATAGGCCCTATGCCCGAAAGTT  3800

3801  AAGGACTACTTTGATAGAGTAGAAAATAGGGGTTCAAACCCCCTTAGTGCCTAGGAAAGAAGGACTTGAACCTCCACTAAAAAGATCAAAACTTTCAGTG  3900
                 ←tRNA-Gln                                                                      NADH 2→
                 tRNA-Met→                                                                        M  N  P  Y
3901  CTTCCATTACACCATCTCCTAGTAAGGTCAGCTGATTAAGCTTTTGGGCCCATACCCCAAATATGTTGGTTAAAATCCTTCCCCTACTAATGAACCCATA  4000
           I  L  S  I  M  L  I  S  L  G  L  G  T  T  L  T  F  A  S  S  N  W  L  L  A  W  M  G  L  E  I  N  T
4001  CATCCTATCCATCATACTTATTAGCCTAGGATTAGGGACAACCCTAACTTTCGCCAGCTCCAACTGACTTCTCGCATGAATGGGATTAGAAATTAACACT  4100
           L  A  I  I  P  L  M  A  N  N  H  H  P  R  A  V  E  A  A  T  K  Y  F  I  T  Q  A  A  A  A  A  L  L
4101  CTAGCCATCATTCCACTTATAGCCAACAATCATCACCCACGAGCAGTAGAAGCTGCAACAAAATACTTCATCACTCAAGCCGCAGCCGCAGCCCTTTTAC  4200
           F  S  S  L  I  N  A  W  Q  S  G  Q  W  M  I  Q  D  M  S  M  P  M  S  A  L  M  T  I  A  I  A  I  K
4201  TATTTTCTAGCCTAATTAACGCCTGACAATCCGGGCAATGAATAATCCAAGACATATCAATACCCATATCTGCACTCATAACCATTGCAATTGCCATTAA  4300
           L  G  V  A  P  V  H  F  W  L  P  E  V  M  Q  G  I  K  L  N  T  G  L  I  L  A  T  W  Q  K  L  A  P
4301  ACTAGGCGTGGCCCCAGTACATTTCTGATTGCCTGAAGTTATACAAGGAATTAAATTAAATACAGGATTAATCTTAGCCACTTGACAAAAACTCGCACCT  4400
           L  A  L  L  Y  Q  I  S  N  N  L  M  P  E  L  M  I  A  L  G  L  M  S  T  I  V  G  G  W  G  G  L  N  Q
4401  CTGGCACTGCTTTATCAAATCTCTAACAACCTAATACCAGAACTCATAATTGCTTTAGGGTTAATATCGACAATCGTGGGGGGATGAGGAGGACTTAACC  4500
           T  Q  I  R  K  I  M  A  Y  S  S  I  A  H  L  G  W  I  I  S  I  M  H  F  M  P  S  L  A  I  I  N  L
4501  AAACACAAATCCGAAAATCATAGCCTACTCATCTATTGCCCATCTAGGCTGAATTATCTCAATCATACACTTCATACCATCCCTCGCCATTATTAACCT  4600
           I  M  Y  I  I  M  T  T  T  M  F  M  I  F  N  T  L  N  S  T  T  I  N  A  L  A  I  N  W  S  K  F  P
4601  AATTATATATATTATTATGACAACAACAATATTCATAATTTTTCAACACCCTAAACTCTACCACAATTAATGCCCTGGCCATTAATTGATCCAAATTCCCA  4700
           A  L  S  A  I  T  M  L  A  L  L  S  L  G  G  L  P  P  L  S  G  F  L  P  K  W  L  I  L  Q  E  L  T  N
4701  GCTCTCTCAGCCATCACCATACTCGCACTCCTATCATTAGGCGGACTCCCACCACTTTCAGGATTCCTGCCAAAATGACTTATTCTTCAAGAACTAACTA  4800
           Q  N  L  A  L  T  A  T  V  M  A  L  S  A  L  L  S  L  Y  F  Y  L  R  L  S  Y  S  L  T  T  T  I  M
4801  ATCAAAATCTCGCACTAACCGCCACAGTAATAGCCCTATCAGCCCTACTTAGCTTATACTTCTATCTCCGACTCTCATATTCACTCACCACAACAATTAT  4900
           P  N  T  Y  Q  H  M  L  N  W  N  I  K  T  K  I  T  F  I  L  P  T  M  M  I  M  T  I  A  M  L  P  I
4901  ACCCAACACATACCAACATATACTTAATTGAAACATCAAAACCAAAATCACCTTTATTCTACCAACAATAATAATTATAACTATCGCCATACTCCCAATC  5000
           S  P  S  I  I  S  M  F      tRNA-Trp→
5001  TCACCCTCAATTATTTCCATATTCTAGGAGTTTAGATTAGCTAAACCAAAAGCCTTCAAAGCTTTAGACAGAGGTTTAAATCCTCTAACTCCTGTTAACT  5100
                                                                              ←tRNA-Ala
5101  AGGATTTGCAGGACTTTATCCCACATCTTATGAATGCAACTCAAACACTTTAATTAAGCTAAAACCCTACTAGATGAGTAGGCCTCGATCCTACATACGT  5200
                                              ←tRNA-Asn
5201  TTAGTTAACAGCTAAAAGCTTTAACCGGCAAGCTTCCACCTAGGCTTTCCTCCCGCCGAAACTCTAAAAGGCGGGAGGAAGCCCCGGCAGGCGTTAACCT  5300
                                              ←tRNA-Cys
5301  ACGTCTCAGGATTTGCAATCCTGCATGCAACTACACTACAAGGCTTGATAAGAGAAGGGATTAAACCTTCCTTTACAGGGCTACAACCTGCCACTTAAAC  5400
                                 COI→
           ←tRNA-Tyr   M  T  I  T  R  W  L  F  S  T  N  H  K  D  I  G  T  L  D  L  I  F  G  A  W  A  G  M
5401  ACTCAGTCATCTTACCTGTGACCATCACCCGCTGACTTTTCTCAACAAACCACAAAGACATTGGCACCCTTGATTTAATCTTTGGTGCCTGAGCCGGAAT  5500
           V  G  T  A  L  S  L  L  I  R  A  E  L  G  Q  P  G  A  L  M  G  D  D  Q  I  Y  N  V  V  G  T  A  H
5501  AGTAGGAACCGCACTAAGCCTCCTAATTCGCGCAGAACTAGGCCAACCAGGAGCCCTAATAGGAGATGACCAAATTTATAATGTTGTCGGCACTGCACAT  5600
           A  F  V  M  I  F  F  M  V  M  P  I  M  I  G  G  F  G  N  W  L  V  P  L  M  I  G  A  P  D  M  A  F  P
5601  GCATTTGTAATAATTTTCTTTATAGTAATGCCAATCATAATCGGAGGATTCGGTAACTGACTTGTACCACTAATAATTGGAGCCCCAGATATGGCCTTCC  5700
           R  M  N  N  M  S  F  W  L  L  P  P  S  L  L  L  L  L  T  S  S  A  V  E  A  G  V  G  T  G  W  T  V
5701  CACGAATAAATAACATAAGTTTCTGATTACTTCCACCATCGCTTCTCCTGCTACTAACCTCTTCTGCAGTAGAGGCCGGTGTCGGAACAGGATGAACTGT  5800
           Y  P  P  L  G  G  N  L  A  H  A  G  A  S  V  D  L  A  I  F  S  L  H  L  V  G  V  S  S  I  L  G  A
5801  ATATCCCCCATTAGGCGGGAATCTAGCACATGCAGGGGCATCAGTTGATTTAGCAATTTTCTCACTCCATTTAGTCGGTGTTTCCTCAATTCTTGGGGCA  5900
           I  N  F  I  T  T  I  I  N  M  K  P  P  S  T  S  Q  Y  Q  T  P  L  F  V  W  S  V  L  V  T  A  V  L  L
5901  ATTAACTTCATTACTACAATTATTAATATGAAACCGCCATCCACCTCACAATACCAAACCCCCCTATTCGTATGATCAGTATTAGTCACTGCAGTCCTCT  6000
           L  L  S  L  P  V  L  A  A  G  I  T  M  L  L  T  D  R  N  L  N  T  T  F  F  D  P  A  G  G  G  D  P
6001  TACTCCTATCCCTTCCAGTATTAGCCGCCGGAATTACAATACTACTAACAGATCGAAATTTAAACACTACCTTCTTTGACCCTGCTGGCGGAGGAGATCC  6100
           I  L  Y  Q  H  L  F  W  F  F  G  H  P  E  V  Y  I  L  I  L  P  G  F  G  M  V  S  H  I  V  A  Y  Y
6101  AATTCTTTACCAACATTTATTCTGATTTTTTGGCCATCCAGAAGTATACATTTTAATTCTTCCAGGATTTGGTATAGTCTCTCATATCGTAGCTTACTAT  6200
           S  G  K  N  E  P  F  G  Y  M  G  M  V  W  A  M  M  A  I  G  L  L  G  F  I  V  W  A  H  H  M  F  T  V
6201  TCAGGTAAAAATGAACCTTTCGGTTATATAGGAATAGTATGAGCAATAATAGCGATTGGACTCCTAGGATTTATCGTATGAGCCCACCACATATTTACAG  6300
           G  M  D  V  D  T  R  A  Y  F  T  S  A  T  M  I  A  I  P  T  G  V  K  V  F  S  W  L  A  T  L  H
6301  TTGGTATAGACGTTGATACTCGAGCTTACTTCACTTCCGCCACAATAATTATTGCTATCCCTACTGGAGTTAAAGTATTTAGCTGACTCGCTACATTACA  6400
           G  G  A  I  K  W  E  T  P  M  L  W  A  L  G  F  I  F  L  L  T  V  G  G  L  T  G  I  I  L  A  N  S
6401  TGGAGGCGCTATTAAATGAGAAACCCCAATACTATGGGCTCTAGGCTTCATCTTCCTATTGACAGTAGGGGGACTAACCGGTATTATTTTAGCAAATTCA  6500
           S  L  D  I  M  L  H  D  T  Y  Y  V  V  A  H  F  H  Y  V  L  S  M  G  A  V  L  A  I  M  G  G  L  V  H
6501  TCACTAGATATTATATTACACGACACATACTACGTGGTAGCACATTTCCATTATGTTTTATCTATAGGAGCTGTCTTGGCCATTATGGGTGGTTTGGTCC  6600
```

FIGURE 3.— Continued

```
            W  F  P  L  F  P  G  Y  T  L  H  P  T  W  T  K  I  H  F  G  V  M  F  I  G  V  N  L  T  F  S  P  Q
6601 ACTGATTCCCCCTATTTCCAGGCTATACACTTCATCCAACCTGAACTAAAATTCATTTCGGAGTAATATTTATTGGGGTCAACTTAACATTCTCCCCACA 6700
        H  F  L  G  L  A  G  M  P  R  R  Y  S  D  Y  P  D  A  Y  T  L  W  N  S  L  S  S  I  G  S  M  I  S
6701 ACACTTCCTAGGACTAGCCGGAATACCCCGACGCTACTCAGATTACCCAGATGCGTACACATTATGAAATTCTTTATCATCCATCGGGTCTATAATTTCC 6800
      L  T  A  V  I  M  F  L  F  I  L  W  E  A  F  A  A  K  R  E  V  Q  T  V  E  L  T  H  T  N  V  E  W  L
6801 CTAACCGCTGTCATTATATTTTTATTTATCCTCTGAGAGGCTTTCGCAGCCAAACGAGAAGTACAAACAGTAGAATTAACCCACACTAATGTTGAATGAC 6900
         H  G  C  P  P  P  Y  H  T  Y  E  E  P  A  F  V  Q  S  P  Q  A  R  E  *
6901 TTCACGGCTGCCCTCCCCCTTACCACACGTACGAAGAACCAGCATTCGTCCAATCCCCACAAGCTCGAGAAAGGAAGGAATTGAACCCCCTTAAATCGGT 7000
                        ←tRNA-Ser(UCN)    tRNA-Asp→
7001 TTCAAGCCGACTGCATAACCATTCTGCCACTTTCTTATGAGATATTAGTTAAACCATAACATTGCCTTGTCAAGGCAAAATTACAGGTTAAACCCCTGTA 7100
          COII→
                M  A  H  P  T  Q  L  G  L  Q  D  A  S  S  P  I  M  E  E  L  L  L  F  H  D  H  A  L  M  T
7101 TATCTTATCTATGGCCCATCCCACACAACTAGGACTACAAGACGCATCCTCTCCTATTATAGAAGAACTACTACTCTTCCACGATCACGCATTAATAACT 7200
     V  F  L  I  S  T  L  V  L  Y  I  I  M  T  A  V  S  T  K  L  T  N  K  H  L  L  D  A  Q  E  I  E  I  V
7201 GTATTTCTAATTAGCACACTTGTACTTTATATTATTATAACCGCAGTTTCAACAAAACTAACCAACAAACATTTACTTGACGCCCAAGAAATTGAGATTG 7300
      W  T  V  M  P  A  L  V  L  I  A  I  A  L  P  S  L  R  I  L  Y  L  M  D  E  I  N  D  P  H  L  T  I
7301 TCTGAACAGTTATACCAGCTCTAGTATTAATTGCTATTGCCCTACCGTCACTTCGAATCCTTTACCTAATAGATGAAATTAACGACCCTCATCTAACTAT 7400
       K  A  T  G  H  Q  W  Y  W  S  Y  E  Y  T  D  Y  D  T  L  N  F  D  S  Y  M  I  P  T  Q  D  L  L  P
7401 TAAAGCCACAGGACATCAATGATACTGAAGCTATGAATATACAGATTACGACACATTAAACTTTGACTCATACATAATTCCCACACAAGATTTATTACCA 7500
      G  Q  F  R  L  L  D  T  D  N  R  M  V  V  P  T  G  S  P  V  R  M  L  I  T  A  E  D  V  L  H  S  W  A
7501 GGACAATTCCGCCTCCTTGACACAGACAATCGAATAGTTGTCCCCACCGGATCTCCTGTGCGCATACTCATTACAGCCGAAGACGTACTTCACTCATGAG 7600
      V  P  S  L  G  L  K  M  D  A  V  P  G  R  L  N  Q  T  T  F  I  A  T  R  P  G  V  F  F  G  Q  C  S
7601 CAGTACCATCATTAGGCCTAAAAATAGATGCGGTGCCAGGACGACTTAATCAAACCACATTCATTGCCACTCGACCAGGAGTTTTCTTCGGTCAATGCTC 7700
       E  I  C  G  A  N  H  S  F  M  P  I  T  I  E  S  A  P  V  K  Y  F  E  S  W  S  S  S  M  L  A  E  S
7701 AGAAATTTGTGGAGCAAACCACAGCTTTATGCCAATTACAATCGAATCAGCGCCTGTTAAATATTTTGAATCTTGATCTTCATCAATACTAGCAGAATCA 7800
     tRNA-Lys→                                                                  ATPase 8→
                                                                                  M  P  Q  L  N  P  N  P  W
7801 TCATTAAGAAGCTAATAGGGCATAGCATTAGCCTTTTAAGCTAAAAATAGGTGACTCCCAACCACCCTTAATGAATGCCACAACTTAATCCAAACCCCTG 7900
      F  T  I  L  I  F  T  W  A  V  F  L  T  I  L  P  N  K  V  T  S  H  K  M  P  N  E  L  L  T  K  D  P
7901 ATTTACCATTTTAATTTTTACCTGAGCTGTATTTCTGACCATTTTACCCAACAAAGTCACCTCACATAAAATACCTAATGAATTATTAACCAAAGACCCT 8000
                        ATPase 6→
                          M  T  L  S  F  L  D  Q  F  A  S  Q  S  F  L  G  I  P  L  I  A  I  A
     S  N  L  L  T  E  I  W  Y  W  P  W  H  *
8001 TCAAACCTCCTTACAGAAATCTGATACTGACCATGACATTAAGCTTTTTGGATCAGTTCGCAAGCCAATCCTTTTTAGGAATTCCACTAATCGCTATCGC 8100
       I  L  I  P  W  M  L  F  P  S  P  Y  K  R  W  M  S  N  R  L  I  T  F  Q  S  W  F  I  A  R  T  T  N
8101 CATCCTAATTCCATGAATATTATTTCCATCCCCCTACAAACGATGAATAAGCAACCGACTAATTACTTTTCAATCCTGATTTATTGCTCGCACCACTAAT 8200
      Q  L  M  L  P  L  N  T  G  A  H  K  W  A  M  I  L  T  A  L  L  L  F  L  M  T  L  N  L  L  G  L  L  P
8201 CAACTTATATTACCACTAAACACTGGTGCACATAAATGAGCTATAATTTTTAACTGCCCTCCTACTATTTCTAATAACTTTAAACCTCCTGGGTCTATTAC 8300
       Y  T  F  T  P  T  T  Q  L  S  M  N  M  A  L  A  V  P  L  W  L  A  T  V  L  I  G  M  R  N  Q  P  T
8301 CATATACATTTACACCAACCACCCAACTATCAATAAATATAGCCCTGGCCGTACCACTATGACTAGCTACCGTATTAATTGGAATACGAAACCAACCAAC 8400
      H  S  L  A  H  L  L  P  E  G  T  P  T  P  L  I  P  I  L  I  I  I  E  T  I  S  L  F  I  R  P  L  A
8401 ACACTCACTAGCCCATCTTCTTCCAGAAGGCACACCGACCCCACTAATCCCCATCCTAATTATTATCGAAACAATTAGCCTCTTTATTCGACCACTCGCT 8500
      L  G  V  R  L  T  A  N  L  T  A  G  H  L  L  I  Q  L  I  S  T  A  T  F  V  M  L  S  I  M  P  T  I  A
8501 CTGGGAGTACGACTAACAGCAAACTTAACCGCAGGCCACTTATTGATTCAACTAATCTCTACAGCAACCTTTGTTATATTATCTATTATGCCAACAATCG 8600
       T  L  T  F  I  V  L  A  L  L  T  I  L  E  I  A  V  A  M  I  Q  A  Y  V  L  V  L  L  L  S  L  Y  L
8601 CCACACTCACATTTATTGTACTAGCTTTACTAACTATTCTAGAAATTGCAGTCGCAATGATTCAAGCTTACGTACTTGTCCTTTTATTAAGCCTGTATCT 8700
      Q  E  N  V      COIII→
                        M  A  H  Q  A  H  A  Y  H  M  V  D  P  S  P  W  P  L  T  G  A  V  A  A  L  L  L  T  S
8701 ACAAGAAAATGTCTAATGGCCCATCAAGCACACGCATATCATATGGTTGACCCAAGCCCATGACCCTTAACAGGGGCAGTTGCCGCTTTACTGCTAACCT 8800
        G  L  A  V  W  F  H  F  K  S  L  T  L  L  A  M  G  L  L  L  M  I  L  T  M  I  Q  W  W  R  D  I  I
8801 CCGGACTAGCAGTATGATTCCATTTCAAATCATTAACCCTACTAGCAATAGGTCTACTATTAATAATTTTAACCATAATTCAATGATGACGAGACATCAT 8900
      R  E  G  T  F  Q  G  H  H  T  P  P  V  Q  K  G  L  R  Y  G  M  I  L  F  I  T  S  E  V  F  F  F  L
8901 TCGTGAGGGAACATTTCAAGGACATCATACCCCACCTGTACAAAAAGGACTCCGCTATGGAATAATCTTATTCATTACATCCGAAGTATTCTTTTTCCTT 9000
      G  F  F  W  A  F  Y  H  S  S  L  A  P  T  P  E  L  G  G  I  W  P  P  T  G  I  T  P  L  D  P  F  E  V
9001 GGGTTTTTCTGAGCCTTTTACCATTCAAGTCTAGCCCCTACACCAGAATTAGGTGGCATTTGACCACCAACCGGTATTACACCTCTAGACCCATTCGAAG 9100
        P  L  L  N  T  A  V  L  L  A  S  G  V  T  W  T  H  H  S  L  M  E  G  K  R  T  E  A  T  Q  A
9101 TTCCTCTTCTTAATACAGCAGTACTACTAGCCTCTGGCGTAACTGTCACATGAACTCACCACAGCCTAATAGAAGGAAAACGAACTGAAGCCACACAAGC 9200
      L  T  L  T  I  L  L  G  L  Y  F  T  A  L  Q  A  M  E  Y  Y  E  A  P  F  T  I  A  D  G  V  Y  G  T
9201 ACTAACCTTAACCATTTTATTAGGCTTATACTTCACCGCCCTTCAAGCAATAGAATATTATGAAGCCCCATTTACTATTGCAGACGGCGTTTACGGAACA 9300
      T  F  F  V  A  T  G  F  H  G  L  H  V  I  I  G  S  T  F  L  A  G  C  L  L  R  Q  I  L  Y  H  F  T  S
9301 ACTTTCTTTGTAGCCACAGGCTTTCATGGACTTCATGTCATCATTGGCTCAACTTTTCTAGCAGGATGCTTATTACGACAAATTCTTTATCACTTTACCT 9400
       S  H  H  F  G  F  E  A  A  A  W  Y  W  H  F  V  D  V  V  W  L  F  L  Y  V  S  I  Y  W  W  G  S
9401 CTTCCCACCACTTTGGCTTTGAAGCTGCTGCATGATATTGACATTTCGTAGACGTAGTCTGACTTTTCCTTTATGTATCAATCTATTGATGAGGATCCTA 9500
```

FIGURE 3. — *Continued*

```
                                                            NADH 3→
      tRNA-Gly→                       ―                     M  N  L  I  L  M  M  I  L  I
9501  ATTTTTCTAGTATTAAAAAGTACAAATGACTTCCAATCATTTAGTCTTGGCTAAATTCCAAGGAAAGATAATGAATTTAATCTTAATAATAATTTTAATT  9600
      S  S  L  I  S  T  I  L  A  I  V  A  F  W  L  P  Q  M  N  P  D  M  E  K  L  S  P  Y  E  C  G  F  D  P
9601  TCCTCCCTAATCTCAACTATCCTAGCTATCGTCGCATTCTGACTCCCGCAAATAAATCCTGATATAGAAAAACTATCACCATACGAATGTGGCTTTGACC  9700
      L  G  S  A  R  L  P  F  S  M  R  F  F  L  V  A  I  L  F  L  L  F  D  L  E  I  A  L  L  L  P  L  P
9701  CGCTCGGATCTGCCCGCCTTCCATTCTCCATACGATTCTTCCTAGTCGCCATTCTATTCCTCCTATTTGACCTAGAAATTGCCCTCCTTCTTCCACTACC  9800
      W  S  T  H  L  D  P  T  L  M  L  M  W  A  F  T  I  I  I  L  L  T  I  G  L  I  Y  E  W  L  Q  G  G
9801  ATGAAGCACCCACTTAGACCCCACTCTTATACTAATATGAGCATTTACCATTATTATTCTCCTGACAATTGGCCTAATTTACGAATGACTTCAAGGTGGA  9900
      L  E  W  A  E                                                           NADH 4L→
             tRNA-Arg→                       ―                               M  T  H  I  M
9901  TTAGAATGAGCAGAATAAGTCTCTAGTCCAACGTAAGATTATTAATTTCGGCTTAATAGATTGTGGCTCAAATCCACAGAGACTTAATGACCCATATCAT  10000
      F  T  F  S  T  A  F  M  L  G  L  S  G  L  T  F  N  R  T  H  L  L  S  A  L  L  C  L  E  G  M  M  L
10001 ATTCACCTTCTCAACCGCCTTTATACTAGGATTATCAGGCCTAACTTTCAATCGCACTCACCTACTCTCTGCCCTGCTATGTTTAGAGGGTATAATATTA  10100
      S  L  F  I  A  L  A  M  W  C  T  Q  N  E  T  M  M  F  S  S  A  P  L  L  L  L  A  L  S  A  C  E  A  G
10101 TCCTTATTTATTGCCCTAGCAATATGATGCACCCAAAATGAGACTATAATATTCTCTTCCGCCCCCCTACTATTACTAGCCCTCTCGGCGTGCGAAGCAG  10200
                                                                  NADH 4→
                                                                  M  L  K  L  L  I  P  T
      L  G  L  S  L  L  V  A  T  A  R  A  H  G  S  D  H  L  Q  N  L  N  L  L  Q  C  *
10201 GCCTCGGCCTAAGCCTACTTGTCGCCACTGCCCGCGCCCACGGGTCTGATCACCTTCAAAAACCTTAACCTCCTACAATGTTAAAACTACTAATTCCAACC  10300
      I  M  L  F  P  M  I  W  T  L  N  P  K  W  L  W  S  A  T  T  T  H  S  L  I  I  A  S  L  S  L  T  L  F
10301 ATTATACTATTCCCAATAATTTGGACACTAAATCCTAAATGATTGTGATCAGCTACCACAACACATAGCTTAATTATTGCTTCCTTATCTCTCACACTAT  10400
      K  C  Y  S  T  T  Q  W  S  N  L  N  Y  M  L  A  T  D  M  I  S  T  P  L  I  I  L  T  C  W  L  L  P
10401 TTAAATGCTACTCCACAACCCAATGATCAAATCTTAATTTATATACTAGCCACCGACATAATCTCCACCCCTTTAATTATCCTAACCTGCTGACTACTTCC  10500
      L  M  I  I  A  S  Q  N  H  M  S  T  E  P  I  N  R  Q  R  S  Y  I  T  L  L  V  S  L  Q  A  L  L  I
10501 ACTAATAATTATTGCTAGCCAAAACCATATATCCACGGAACCCATCAACCGACAACGAAGTTACATTACACTACTAGTATCCCTGCAAGCCCTACTTATT  10600
      M  A  F  S  A  T  E  I  I  L  F  Y  I  M  F  E  S  T  L  I  P  T  L  I  I  T  R  W  G  N  Q  T  E
10601 ATAGCCTTCAGCGCCCACAGAAATTATTTTATTTTATATTATATTTGAATCCACCTTAATCCCTACCCTTATTATTATTACACGCTGAGGAAACCAAACAG  10700
      R  L  N  A  G  I  Y  F  L  F  Y  T  L  A  G  S  L  P  L  L  V  A  L  L  Y  L  N  T  A  G  S  L
10701 AACGTCTTAACGCTGGCATCTACTTCCTATTTTATACTCTAGCCGGCTCACTTCCATTACTAGTTGCCCTCCTATATTTATATAACACAGCCGGCTCGCT  10800
      S  F  I  S  M  N  L  I  S  I  P  P  N  T  W  T  N  T  F  L  W  V  A  C  V  T  A  F  L  V  K  M  P
10801 GTCATTTATTTCAATAAACTTAATCTCTATCCCCCCTAACACTTGAACCAACACTTTCTTATGAGTAGCCTGCGTAACCGCCTTCCTAGTAAAAATACCT  10900
      L  Y  G  V  H  L  W  L  P  K  A  H  V  E  A  P  V  A  G  S  M  I  L  A  A  I  L  L  K  L  G  G  Y  G
10901 TTATACGGAGTTCATCTCTGACTCCCCAAAGCCCATGTAGAAGCCCCAGTTGCCGGATCAATAATTCTTGCGCCATCCTACTAAAATTGGGCGGATACG  11000
      M  I  R  M  T  I  M  L  E  P  A  T  K  S  L  A  Y  P  F  I  I  L  A  L  W  G  I  I  M  T  G  S  I
11001 GAATAATTCGAATAACCATTATACTAGAACCTGCAACCAAATCATTAGCCTACCCATTCATTATTCTTGCATTATGAGGAATCATTATGACAGGGTCAAT  11100
      C  M  R  Q  S  D  M  K  S  L  I  A  Y  S  S  V  S  H  M  G  L  V  A  S  G  I  L  I  Q  T  T  W  G
11101 TTGTATACGACAATCAGACATAAAATCCCTAATCGCCTACTCCTCAGTCAGCCATATAGGACTAGTTGCCTCCGGCATCCTAATCCAAACCACATGGAGGT  11200
      F  T  G  A  I  I  L  M  I  A  H  G  L  T  S  S  A  L  F  C  L  A  N  T  A  Y  E  R  T  H  S  R  T  L
11201 TTTACAGGAGCAATTATCTTAATAATTGCCCACGGGTTAACATCATCAGCCCTATTCTGCCTAGCCAACACGGCCTACGAACGAACACACTCCCGAACTC  11300
      L  L  A  R  G  M  Q  I  I  L  P  L  M  A  T  W  W  F  I  M  S  L  A  N  M  A  L  P  P  L  P  N  L
11301 TCCTCCTAGCACGAGGAATACAAATCATCTTACCATTAATAGCGACTTGATGATTTATTATAAGTCTAGCAAACATAGCTTTACCTCCACTTCCCAACCT  11400
      M  G  E  L  M  I  L  V  S  M  F  N  W  S  N  W  T  I  L  L  T  G  T  G  T  L  I  T  A  S  Y  S  L
11401 AATAGGAGAACTAATAATCCTAGTATCTATATTTAAACTGATCCAACTGAACTATTTTATTAACAGGTACAGGAACCCTAATTACAGCTAGTTACTCACTA  11500
      Y  L  Y  M  S  S  Q  R  G  P  T  P  N  N  L  T  F  M  E  L  S  H  T  R  E  H  L  L  L  T  L  H  I  I
11501 TATCTTTATATATCATCTCAACGAGGACCAACCCCCAACAACCTAACCTTTATAGAACTTTCCCACACCCGAGAACACCTCCTACTAACCTTACATATTA  11600
      P  I  I  L  L  M  I  K  P  E  L  I  W  G  W  C  W     tRNA-His→
11601 TTCCAATTATCCTCCTAATAATTAAACCTGAACTTATCTGAGGATGATGTTGATGTTAATATAGTCTAAATAAGACATTAGATTGTGATTCTAATAATAG  11700
         tRNA-Ser(AGY) →                 ―                                              tRNA-Leu
11701 AAGTTAAACTCTTCTTATTGACCGAGAGATGCTCGCGCACCAAGAACTGCTAATTCTTATTACCTTGGTTAAACTCCAAGGATCACTCGATGCTTTCAAA  11800
                                                            NADH-5→
      (CUN)→                        ―                       M  S  I  S  Q  L  S  Q  M  F  M  T  C
11801 GGATAATAGATATCCATTGGTCTTAGGAACCAAAAACTCTTGGTGCAAATCCAAGTGAAAGCTATGTCTATCTCCCAATTATCACAAATATTTATAAACAT  11900
      L  S  L  T  M  I  I  L  I  L  P  I  T  F  S  F  I  T  K  P  S  N  K  W  P  F  Q  V  K  N  A  V  K
11901 GCCTCTCCCTAACAATAATTATTCTCATCCTACCAATCACGTTTTCCTTCATTACGAAACCATCAAACAAATGACCTTTCCAAGTCAAAAATGCTGTAAA  12000
      L  S  F  F  V  S  L  I  P  S  I  T  C  L  N  L  N  L  Q  S  F  T  I  Y  Y  Q  W  F  S  I  S  S  T
12001 ATTATCCTTCTTTGTAAGCCTAATTCCATCAATTACTTGCTTAAACTTAAATCTCCAATCCTTTACCATCTACTACCAATGATTTTCTATTTCATCCACA  12100
      K  I  N  I  S  L  Q  F  D  Q  Y  S  M  I  F  M  T  I  A  L  Y  V  T  W  S  I  L  E  F  A  I  Y  Y  M
12101 AAAATTAACATTAGCCTACAATTTGATCAATACTCCATAATTTTTATAACAATTGCACTCTACGTAACCTGATCAATTCTAGAATTCGCTATTTATTACA  12200
      H  T  D  I  L  I  N  R  F  F  K  Y  L  L  T  F  L  I  A  M  M  I  L  V  T  A  N  N  M  F  Q  L  F
12201 TACACACTGACATTTTAATTAACCGATTCTTTAAATACCTATTAACATTTCTAATTGCCATAATAATTCTGGTCACAGCTAATAACATATTCCAACTTTT  12300
      I  G  W  E  G  V  G  I  M  S  F  L  L  I  G  W  W  Y  G  R  A  D  A  N  M  A  A  L  Q  A  V  I  Y
12301 TATCGGCTGAGAAGGAGTAGGAATCATATCATTCCTACTAATTGGGTGATGATATGGCCGCGCCGATGCCAACATAGCCGCATTACAAGCTGTAATTTAT  12400
```

FIGURE 3.— Continued

```
               N  R  V  G  D  I  G  L  M  M  T  M  S  W  L  L  I  N  T  N  S  W  D  I  Q  Q  L  F  G  L  T  K  N  M
12401  AATCGTGTAGGAGACATCGGTCTCATAATAACAATATCTTGACTTTTAATCAACACTAACTCATGAGATATTCAACAACTATTTGGCCTTACAAAAAATA  12500
               D  T  T  L  P  A  T  G  L  L  L  A  A  T  G  K  S  A  Q  F  G  L  H  P  W  L  P  A  A  M  E  G  P
12501  TAGACACAACACTTCCTGCAACCGGACTGCTACTAGCAGCAACAGGCAAATCCGCCCAATTTGGCCTTCACCCATGACTACCCGCAGCAATAGAAGGTCC  12600
               T  P  V  S  A  L  L  H  S  S  T  M  V  V  A  G  I  F  L  L  I  R  L  H  P  L  I  E  N  N  N  N  I
12601  AACTCCAGTCTCCGCCCTACTTCACTCCAGCACAATAGTTGTTGCCGGAATCTTCCTGTTAATCCGCCTTCACCCACTCATTGAAAACAATAACAATATC  12700
               L  T  A  A  L  C  L  G  A  I  T  T  L  F  T  A  T  C  A  L  T  Q  N  D  I  K  K  I  V  G  F  S  T  S
12701  CTTACCGCTGCACTCTGCCTAGGCGCAATTACTACCCTCTTCACCGCCACCTGCGCCCTGACACAGAATGACATCAAAAAAAATCGTTGGATTCTCTACAT  12800
               S  Q  L  G  L  M  M  V  A  I  G  L  N  Q  P  Q  L  A  F  L  H  I  C  T  H  A  F  F  K  A  M  L  F
12801  CCAGTCAACTAGGCCTAATAATAGTAGCAATTGGACTTAACCAACCCCAATTAGCATTTCTTCACATCTGCACCCACGCATTTTTCAAAGCCATACTCTT  12900
               L  C  S  G  S  I  I  H  S  L  N  D  E  Q  D  I  R  K  M  G  G  I  N  K  T  L  P  L  T  S  S  C  L
12901  CCTATGTTCTGGATCAATTATCCACTCCTTGAATGACGAACAAGACATCCGTAAAATAGGCGGAATTAATAAAACACTCCCATTAACCTCCTCATGCTTA  13000
               T  I  G  S  L  A  L  M  G  T  P  F  L  A  G  F  F  S  K  D  A  I  I  E  A  I  N  T  S  H  L  N  A  W
13001  ACTATCGGCAGCCTCGCACTGATAGGAACACCATTTCTTGCGGGCTTCTTCTCAAAAGATGCAATCATCGAAGCCATTAACACCTCACACCTAAACGCCT  13100
               A  L  V  L  T  L  I  A  T  S  F  T  A  V  Y  S  L  R  I  I  Y  F  V  L  M  N  H  P  R  T  L  P  L
13101  GAGCCCTAGTACTAACACTAATCGCCACCTCATTCACCGCCGTGTACAGCTTACGCATCATCTATTTTGTCTTAATAAATCATCCTCGTACCCTCCCCCT  13200
               S  P  V  N  E  N  N  P  L  I  A  N  P  I  K  R  L  A  W  G  S  I  I  A  G  L  I  L  C  Q  Y  I  L
13201  CTCTCCCGTAAATGAAAATAATCCCTTAATTGCTAATCCAATTAAACGCCTTGCATGAGGAAGCATTATTGCCGGATTAATTCTTTGTCAATATATTCTT  13300
               P  N  K  T  Q  T  L  T  M  T  P  M  L  K  L  T  A  L  I  V  S  L  L  G  L  L  T  A  L  E  L  A  S  M
13301  CCTAATAAAACCCAAACACTTACAATAACACCGATACTAAAACTAACTGCCCTTATTGTATCATTATTAGGACTACTTACAGCATTAGAACTGGCCTCCA  13400
               A  N  K  Q  I  K  I  N  P  T  K  F  T  H  N  F  S  N  M  L  G  F  Y  P  H  I  M  H  R  L  M  S  K
13401  TAGCAAATAAACAAATCAAAATTAATCCTACAAAATTTACCCATAACTTCTCCAACATACTGGGTTTCTATCCTCATATTATACACCGCTTAATATCTAA  13500
               L  P  L  M  L  G  Q  I  S  A  T  Q  M  S  D  Q  L  W  M  E  K  L  G  P  K  G  I  A  H  T  Q  L  L
13501  ATTACCGCTAATACTAGGACAAATCTCTGCTACACAAATATCAGACCAACTATGAATAGAGAAATTAGGACCAAAAGGCATTGCCCACACACAACTTCTC  13600
               V  T  Q  K  I  T  H  V  H  K  G  L  I  K  T  Y  L  S  I  M  M  L  S  I  I  I  I  T  I  I  M  I  T
13601  GTTACCCAAAAAATTACACATGTACATAAAGGCCTAATTAAAAACATACTTATCAATTATAATACTATCAATCATTATTATCACAATTATCATTATAATTA  13700
                 *
                 *  V  A  R  L  C  G  R  S  Y  G  R  T  L  E  L  V  V  L  L  A  M  L  L  V  W  G  A  I  F  L  M  Y
13701  CCTAACCGCACGTAAACAACCACGAGAATATCCTCGAGTTAATTCTAGCACAACTAATAAAGCTATTAATAATACCCACCCAGCAATAAACAACATATAA  13800
               G  G  A  N  Y  V  E  A  V  G  V  Y  R  R  M  A  G  F  Y  E  E  V  D  D  F  L  E  M  D  G  V  F  A
13801  CCTCCAGCATTATATACCTCTGCTACCCCAACATATCGCCGCATGGCCCCAAAATATTCCTCAACATCATCAAATAATTCTATATCCCCAACAAATGCTA  13900
               V  W  A  V  I  V  L  F  M  Y  V  L  V  Y  S  F  V  E  W  S  G  W  A  N  P  Y  P  E  A  A  L  A  P  S
13901  CCCATGCTACAATAACTAAAAACATATATACTAATACATAAGAAAAAACCTCCCAAGAACCCCACGCATTAGGATAAGGCTCCGCTGCCAATGCAGGAGA  14000
               Y  A  F  V  V  L  M  G  G  L  Y  I  L  F  L  I  M  S  L  F  T  M  G  C  Q  M  L  M  G  C  G  V  G
14001  ATAAGCAAATACTACTAACATCCCTCCCAAATAAAATTAAAAAAAAGAATTATAGATAAAAAAGTTATACCACATTGCATTAACATTCCACACCCAACTCCC  14100
                                                                                                      ←NADH 6
               A  G  V  M  L  G  F  A  A  F  Y  P  S  P  N  S  A  V  A  I  L  S  I  L  F  M  V  S  F  V  V  M  L
14101  GCTCCAACTATCAACCCAAAAGCTGCAAAATAAGGAGAAGGATTAGATGCTACTGCAATTAAACTAATTAAAAATATTACAGAAAAAACTACCATAAGTA  14200
                                                                              Cyt b→
               V  M                               ___                    ←tRNA-Glu  M  A  I  I  R  K  T  H
14201  CCATAATTCTTGCCAGGACTTTAACCAGGACTAATGACTTGAAAAATCACCGTTGTAACTCAACTACAAAAATCAAATGGCAATCATCCGCAAAACCCAC  14300
               P  L  A  K  I  I  N  S  A  F  I  D  L  P  A  P  S  N  I  S  S  W  W  N  M  G  S  L  L  G  L  C  L  I
14301  CCTCTAGCAAAAATTATTAACAGTGCATTCATTGACCTACCAGCCCCATCAAACATCTCATCATGATGAAACATAGGCTCCCTACTTGGATTATGCCTAA  14400
               A  Q  I  I  T  G  L  F  L  A  M  H  Y  V  S  D  I  N  S  A  F  S  S  V  A  H  I  C  R  D  V  N  Y
14401  TCGCACAGATTATCACAGGACTATTCTTAGCCATACATTACGTTTCCGATATTAACTCAGCTTTCTCCTCAGTCGCACACATCTGCCGTGATGTAAACTA  14500
               G  W  L  I  R  N  F  H  A  N  G  A  S  L  F  F  I  C  I  Y  L  H  I  A  R  G  L  Y  Y  G  S  Y  L
14501  TGGATGACTAATCCGAAACTTCCACGCAAACGGAGCATCCCTATTCTTCATTTGCATTTACTTACATATCGCGCGCGGTCTATATTATGGATCCTACCTC  14600
               Y  T  E  T  W  N  M  G  V  I  L  L  L  L  T  M  M  T  A  F  V  G  Y  V  L  P  W  G  Q  M  S  F  W  G
14601  TACACAGAAACCTGAAACATAGGGGTAATTCTACTCCTCCTCACCATAATAACCGCATTTGTAGGTTATGTCCTTCCATGAGGGCAAATATCCTTCTGAG  14700
               A  T  V  I  T  N  L  L  S  A  I  P  Y  I  G  D  T  L  V  Q  W  I  W  G  G  F  S  V  D  K  P  T  L
14701  GTGCCCAGTTATTACCAATCTTCTATCAGCAATCCCATACATCGGAGATACCTTAGTACAATGAATCTGAGGAGGGTTTTCAGTTGATAAACCAACACT  14800
               V  G  I  N  S  D  A  D  Q  I  P  F  H  P  Y  F  T  F  K  D  L  L  G  F  I  I  L  L  L  I  I  M  L
14801  TACCCGATTCTTTGCATTCCACTTCATTCTTCCCTTCGCAATCGCAGCAGCGTCCCTCGTCCATATTGTATTTCTTCATGAAACAGGGTCAAACAACCCA  14900
               A  L  L  S  P  N  L  L  N  D  P  G  N  F  T  P  A  N  P  L  I  T  P  P  H  I  K  P  E  W  Y  F  L
14901  GTAGGAATTAATTCCGATGCAGACCAAATTCCATTTCACCCTTACTTTACCTTTAAAGACCTACTAGGCTTCATTATCTTACTACTAATTATTATTATAT  15000
               A  L  L  S  P  N  L  L  N  D  P  G  N  F  T  P  A  N  P  L  I  T  P  P  H  I  K  P  E  W  Y  F  L
15001  TAGCATTACTTTCACCCAACCTATTGAATGACCCAGGAAATTTTACTCCAGCCAACCCCTTAATTACCCCCCCTCATATTAAACCTGAATGATATTTCCT  15100
               F  A  Y  A  I  L  R  S  I  P  N  K  L  G  G  V  L  A  L  L  F  S  I  L  I  L  M  L  V  P  L  L  H
15101  ATTTGCCTACGCAATTCTACGTTCTATTCCCAACAAACTAGGAGGAGTGTTAGCCCTACTATTCTCCATCCTTATCCTAATGTTAGTACCATTACTCCAC  15200
               T  S  K  I  R  S  A  T  F  R  P  L  F  K  I  T  L  W  I  L  A  A  D  V  L  I  L  T  W  I  G  G  Q  P
15201  ACCTCAAAAATTCGCAGCGCCACATTCCGCCCGCTATTCAAAATTACACTCTGAATTCTCGCAGCAGACGTCCTCATCTTAACATGAATTGGAGGACAAC  15300
               V  E  D  P  Y  I  I  I  G  Q  A  A  S  I  L  Y  F  L  I  F  L  V  L  M  P  L  S  G  W  L  E  N  K
15301  CAGTAGAAGACCCGTACATTATTATTGGCCAAGCCGCCTCGATCCTTTACTTCTTAATTTTCTTAGTACTTATACCACTATCAGGTTGACTAGAAAACAA  15400
```

FIGURE 3.—*Continued*

```
       M   L   N   R   D    tRNA-Thr→
15401  AATACTCAACCGCGATTGCCCTAGTAACTTAAAGATAAAGTGTCGGTCTTGTAAACCGAACATGATGGTTAAACTCCATCCTTGGGCTCAAAGAGAAGAG  15500
                                                 ←tRNA-ProControl Region→                        TAS-1
15501  ACTTTAACTCCTGCCTTTAACTCCCAAAGCTAAAATTCTACTTAAACTACTCTTTGATCATGTATGTATATAAGACACAATATTTAAACAAACATTACAT  15600
       ____   ___TAS-2___
15601  TATCCCTACATAAACCATATAAACTATTAATCAGGTATTACACCCACCATATCAAAGTACACTACCCTGGTCACTACTACCCCACCTCCCCCCCAAAAT   15700
       ___TAS-3___                                       ___TAS-4___
15701  ATTACATACTATTAACTATATTAGATTTATATGTATATATTACATTAAATTACTTTCAATATAAACCATTAACTTATATATCCCAATAATTAATAACAA  15800
15801  CTAATATGCTCCATGAGCTCCATGAACAACAATTAATCAAGAAATACCTTACTAAAACCACAACATGGATACTAACTACAGCCTTATAAGTACATATATA  15900
15901  TACTTAACACCTATTAAGCCAGGAATATACATGTAATTTACCATTAAATTATGTACAGCATTCAGGAGAGTCATTCGTTTTCATATTTATCAACAATATT  16000
16001  CCATAATAACAGTCTCGTCCGGATTGTAACCACCATCCTAAACGGTTCCACCCTACCCGAACCATCATTATAATCTAAGACTTATTATCCGTGATACACT  16100
                          ___CB-D___
16101  AACGTTCCATTCCAGAACATTTGGTTCCTAATTCAGGCACATAAATAGATAATTCCCTATACCTGATATTATCGAGGCTATAGTGACTGTACGGAACATG  16200
16201  GGTCTCCTGGCACGCGACATGGTTTAAACAGTCCTCGTACGTTCCGCACTGCGGTTGTTTAACCTGATCTGCCTCATTAGCCTTTTTCCTCCTTCACTTT  16300
                                                                                          CSB-I
16301  CATCACCTTGCCACGACTGGTTATTCAGATTTTTAGTCAAGGCCCGGTACATTTACTTTGCTTAAGACTTGATCATTCAACCTTGAATGACATACAATTC  16400
16401  TTTTACCATTAATTATTCTCTCAAGAGCATAACTACTTTTACATCCCTTGGTGCTTTCTTAATAAATTAAAATTTTAAAACGCGAAAATAGCCTAAAACC  16500
       ___CSB-II___                                        ___CSB-III___
16501  CCCTACCCCCTTAAAAAAGACGTTTTTGGACTTTTTGGCCCCGTAAAACCCCGAAAACAGGAAAGTGCCTAGAAACGCAAGAGAATTGTGTTGCTGCAAA  16600
16601  TCATCTTGCAGTGTTGCACTACGA      16624
```

FIGURE 3.—*Continued*

than to those of either the lungfish or the lamprey (Table 4).

All protein encoding genes in the bichir mitochondrial genome use the initiation codon ATG except for *COI*, which uses GTG. The use of GTG for initiation of translation in *COI* is shared with other fish mitochondrial genomes that have been sequenced completely and with the chicken mitochondrial genome. Three genes (*ATPase 8, ND4L*, and *ND5*) use TAA as stop codon, eight genes (*ND1, ND2, COII, ATPase 6, COIII, ND3, ND4*, and *cyt b*) use the incomplete stop codon T, and the remaining two genes (*COI* and *ND6*) use an AGR stop codon (Table 3). Of the complete vertebrate genomes sequenced, no ray-finned fish uses AGR as a stop codon, whereas lamprey, lungfish, frog, chicken and mammals do use AGR stop codons.

The overall codon usage of the 13 protein genes of the bichir mtDNA genome shows a strong bias against guanine at third positions. The anti-G bias in bichir is the second strongest (4.4%) after lamprey (3.8%), while other fish and frog are less strongly biased (MEYER 1993) (Table 5). As seen in other mitochondrial genomes (NAYLOR *et al.* 1995) pyrimidines are overrepresented compared with purines in the second codon position (percent C+T = 68.3) due to the hydrophobic character of proteins encoded by mitochondrial genes.

Transfer RNA genes: The bichir mitochondrial genome contains 22 transfer RNA genes that are interspersed between ribosomal RNA and protein coding genes. The transfer RNA genes range in size from 67 to 75 nucleotides. All bichir tRNA genes can be folded into conventional cloverleaf secondary structures. As seen in other vertebrate tRNAs formation of G-U, and other atypical pairings are found in the stem regions. All proposed cloverleaf structures, except that of tRNA$^{Ser(AGY)}$ contain 7 bp in the amino acid stem, 5 bp

in the TΨC stem, 5 bp in the anticodon stem and 4 bp in the DHU stem. In most vertebrates tRNA$^{Ser(AGY)}$ has a reduced DHU arm. The bichir tRNA$^{Ser(AGY)}$, similar to that of the lamprey tRNA$^{Ser(AGY)}$, has no recognizable DHU stem and loop. Rainbow trout, carp and cod on the other hand posses a tRNA$^{Ser(AGY)}$ which is much less truncated (ZARDOYA *et al.* 1995).

Ribosomal RNA genes: The 12S and 16S ribosomal RNA genes in bichir are 950 and 1655 nucleotides long, respectively. They are situated as in other vertebrates between tRNA^{Phe} and tRNA$^{Leu(UUR)}$ and are separated by the tRNA^{Val}. The ribosomal RNA genes follow the same trend as the protein encoding genes and show greater sequence similarity to ray-finned fish than to lamprey or lungfish (Table 4).

Noncoding regions: The control region, which is the largest noncoding region found in the bichir mtDNA, is 1068 nucleotides long and is situated between tRNA^{Pro} and tRNA^{Phe} genes. The control region of vertebrates contains the site of initiation for heavy strand replication as well as promoters for the transcription of both the light strand and the heavy strands. The bichir control region has an overall base composition that is rich in A and T (percent A+T = 64.3). Between positions 16,100 and 16,300 (Figure 3), a G-rich region is found within which the central conserved block D (SOUTHERN *et al.* 1988) was identified. This sequence block is conserved between bony fish and mammals (LEE *et al.* 1995), but has not been found in the lamprey control region (LEE and KOCHER 1995). The conserved sequence blocks CSB-II and CSB-III, which are thought to be involved in the generation of the RNA primer needed for initiation of heavy strand replication (WAL-BERG and CLAYTON 1981), are present in the right domain of the bichir control region. Similar to lungfish (ZARDOYA and MEYER 1996a) and frog (ROE *et al.* 1985),

TABLE 3

Localization of features in the mitochondrial genome of
Polypterus ornatipinnis

Feature	From	To	Size, bp	Codon Start	Codon Stop
tRNA-Phe	1	71	71		
12S rRNA	72	1021	950		
tRNA-Val	1022	1092	71		
16S rRNA	1093	2747	1655		
tRNA-leu(UUR)	2748	2822	75		
NADH 1	2823	3780	958	ATG	T–
tRNA-Ile	3781	3853	73		
tRNA-Gln	3854	3920	67 (L)		
tRNA-Met	3921	3989	69		
NADH 2	3990	5025	1036	ATG	T–
tRNA-Trp	5026	5094	69		
tRNA-Ala	5100	5168	69 (L)		
tRNA-Asn	5170	5242	73 (L)		
tRNA-Cys	5278	5346	69 (L)		
tRNA-Tyr	5347	5416	70 (L)		
CO I	5418	6974	1557	GTG	AGG
tRNA-Ser (UCN)	6966	7036	71 (L)		
tRNA-Asp	7039	7107	69		
CO II	7111	7801	688	ATG	T–
tRNA-Lys	7802	7874	73		
ATPase 8	7875	8042	168	ATG	TAA
ATPase 6	8033	8714	682	ATG	T–
CO III	8716	9499	784	ATG	T–
tRNA-Gly	9500	9570	71		
NADH 3	9571	9916	346	ATG	T–
tRNA-Arg	9917	9986	70		
NADH 4L	9987	10283	297	ATG	TAA
NADH 4	10277	11654	1378	ATG	T–
tRNA-His	11655	11723	69		
tRNA-Ser (AGY)	11724	11789	66		
tRNA-Leu (CUN)	11792	11863	72		
NADH 5	11864	13705	1842	ATG	TAA
NADH 6	13701	14204	504 (L)	ATG	AGG
tRNA-Glu	14205	14273	69 (L)		
CYT b	14277	15417	1141	ATG	T–
tRNA-Thr	15418	15488	71		
tRNA-Pro	15489	15556	68 (L)		
Control region	15557	16624	1068		

a putative CSB-I with a reduced motif of GACAT can be identified at position 16,398 (Figure 3).

At the 3′ end of the control region, the palindromic sequence motives TACAT and ATGTA are repeated four and three times, respectively. These motives have also been found at the 3′ end of other vertebrate control regions and are thought to act as a recognition site for the termination of heavy strand elongation by forming stable hairpin-loop structures (SACCONE *et al.* 1991). In the left domain, we could also identify four putative termination associated sequences (TASs) based on the consensus mouse TAS sequence proposed by DODA *et al.* (1981) (Figure 3).

The origin of light strand replication (O_L) is located between the $tRNA^{Asn}$ and $tRNA^{Cys}$ genes and is 46 nucleotides in length (Figure 3). The sequence has the potential to form a stable stem-loop structure with 11 bp in the stem and 12 nucleotides in the loop. The conserved motif 5′-GCCGG-3′ at the base of the stem within the $tRNA^{Cys}$ gene, which has been shown to be involved in the transition from RNA synthesis to DNA synthesis (HIXSON *et al.* 1986), is also conserved in the bichir. The most interesting feature of the bichir O_L is a stretch of T-rich sequence in the loop. This stretch of T-rich sequence in humans has been shown to be involved in the synthesis of a RNA primer for initiation of light strand replication (WONG and CLAYTON 1985). All mammalian mtDNAs sequenced so far have this T-rich sequence, whereas all other fish mtDNAs sequenced have a stretch of cytosines or, as is the case for lungfish, a stretch of cytosines and thymines (ZARDOYA and MEYER 1996a).

Phylogenetic analyses of bichir mitochondrial genome: To phylogenetically correctly place the bichir among vertebrates, and especially to assess its relationship to ray-finned and lobe-finned fish, we used the mitochondrial genes of the following species for phylogenetic analyses: human (ANDERSON *et al.* 1981), whale (ARNASON and GULLBERG 1993), marsupial (JANKE *et al.* 1994), chicken (DESJARDINS and MORAIS 1990), frog (ROE *et al.* 1985), lungfish (ZARDOYA and MEYER 1996a), rainbow trout (ZARDOYA *et al.* 1995), carp (CHANG *et al.* 1994), loach (TZENG *et al.* 1992) and lamprey (LEE and KOCHER 1995). 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

TABLE 4

Percentage nucleotide similarity of bichir mitochondrial genes against their homologues of other available fish

	12S	16S	ND1	ND2	ND3	ND4	ND4L	ND5	ND6	ATP6	ATP8	COI	COII	COIII	CYT b
bichir-lamprey	68.3	66.8	66.8	57.7	68.4	61.9	57.4	61	56.9	63.9	50.6	77.6	71	74.4	65.6
bichir-ray-finned fish[a]	73.9	71.2	71	63.7	70.5	69.4	67.7	67.3	61	70.9	61.7	78.5	75.3	77.1	72.2
bichir-lungfish	71.2	68.9	67.4	61.9	66.2	65	58.6	62.9	59.3	66.7	57.1	76.2	73.8	75.8	69.3

[a] Ray-finned fish refers to the average sequence similarity between bichir against trout, carp and loach.

TABLE 5

Base composition for the 13 protein-coding genes of fish and an amphibian

	Codon position	A	G	C	T
Lamprey	1	30.4	22.6	22.9	24.1
	2	19	12.9	26.5	41.6
	3	41.3	3.8	21.5	33.4
Polypterus	1	29.9	23	23.7	23.4
	2	18.9	12.8	27	41.3
	3	42.7	4.4	27.8	25.1
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	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
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	23.3	25.8
	2	20.5	11.6	27.2	40.7
	3	41.2	6.5	22.3	30

Values are expressed as percentage.

genes were excluded from the phylogenetic analyses. The analyses were performed using nucleotide sequences and third codon positions were excluded since they are largely saturated with back mutations at this level of phylogenetic inquiry (e.g., CAO et al. 1994; RUSSO et al. 1996; ZARDOYA and MEYER 1996a). Similarly, all tRNA genes were aligned taking their secondary structures into account. DHU and TΨC arms were omitted due to ambiguity in alignments, and all positions were weighted equally. Moreover, both rRNA genes were aligned also taking their secondary structures into account and questionably aligned sequences were omitted from the analyses. For rRNA genes, all positions were weighted with a transition/transversion ratio of 2:1 to account for the faster rate for transitions over transversions in these genes (e.g., MEYER 1993; ORTÍ and MEYER 1996). In all analyses, gaps in alignments were treated as missing data.

In their investigation of lungfish relationships among vertebrates, ZARDOYA and MEYER (1996a) found that if the lamprey is used as outgroup, vertebrate trees with unexpected topologies and low bootstrap values are found regardless of which phylogenetic method is used. The same results are obtained when bichir is added to the analysis and lamprey is used as outgroup. It seems that too many multiple substitutions might have accumulated along the fast evolving lamprey lineage, hindering the recovery of correct phylogenetic relationships among vertebrates (RUSSO et al. 1996; ZARDOYA and MEYER 1996a,c). We used the same taxa that were included in the ZARDOYA and MEYER (1996a) study to
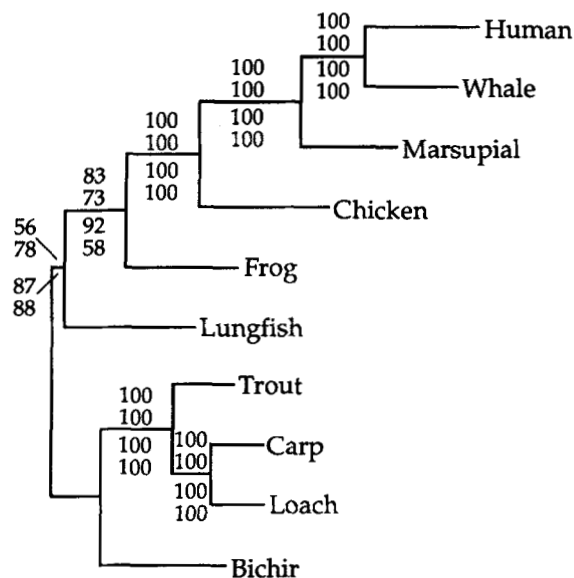
FIGURE 4.—Fifty percent majority rule bootstrap (FELSENSTEIN 1985) consensus tree of jawed vertebrates based on 1000 replications. Two data sets were subjected to MP (bootstrap values above branches) and NJ (bootstrap values below branches) analyses. The first data set comprises a combination of all mitochondrial genes (bootstrap values upper of each pair of numbers). The second data set includes all mitochondrial protein coding genes combined (bootstrap values lower of each pair of numbers). ML analyses with the same two data sets yielded the same topology. Bichir, rainbow trout, carp, and loach were used as outgroup for all analyses.

investigate the performance of the bichir as outgroup in resolving vertebrate phylogenetic relationships. The phylogenetic analyses were performed using a data set of all mitochondrial genes combined and individual data sets of combined protein coding genes, tRNAs or rRNAs only. All three commonly used phylogenetic methods (MP, NJ, ML) arrived at identical and strongly supported topologies when bichir, rainbow trout, carp and loach were used as outgroup taxa and the combined set of all mitochondrial genes was analyzed (Figure 4).

With the separate protein data set, the expected phylogeny was also obtained with all three phylogenetic methods used (Figure 4). Strikingly, the separate data sets for tRNAs or rRNAs were unable to resolve the expected vertebrate relationships. In the case of the rRNA data set, the MP and ML analyses placed frog and lungfish as the sistergroup of teleosts and this grouping was supported by a bootstrap value of 56% (MP). NJ analysis of this data set arrived at a tree that placed frog and lungfish into a monophyletic group basal to tetrapods, this clade was supported by a bootstrap value of 52%. For the tRNA data set, all three phylogenetic analyses (MP, NJ, and ML) arrived at unexpected trees in which the lungfish was basal to teleosts and tetrapods. This topology was supported with bootstrap values of 74 and 57% for MP and NJ analyses, respectively. In the parsimony analyses of the tRNA and

TABLE 6

Statistical confidence of maximum likelihood trees

	tRNAs			rRNAs		
Tree topology	log l	Δli	SE	log l	Δli	SE
1. (bichir,(trout,(carp,loach)),(lungfish,(frog,(chicken,(marsupial,(whale,human))))))	−8746	−6.7	6.3	−11,004	−11.2	16.6
2. (bichir,((frog,lungfish),(trout,(carp,loach))),(chicken,(marsupial,(whale,human))))	−8762	−22.6	11	−10,992	—	—
3. (bichir,lungfish,((trout,(carp,loach)),(frog,(chicken,(marsupial,(whale,human))))))	−8739	—	—	−11,012	−19.4	17.1

Δli represents the difference in log-likelihood between the best tree (i) obtained for each data set and the expected tree (tree #1). SE is the standard error. Trees are declared significantly different when the difference in log-likelihood is larger than 1.96 standard error.

rRNA data sets, only a few more steps are needed to recover the expected topology (Figure 4), suggesting that the shortest trees obtained for those two data sets are poorly supported and not statistically different from the expected tree. 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 data set and that of the correct tree were calculated by the formula of KISHINO and HASEGAWA (1989). Both the rRNA data set and the tRNA data set exhibited log-likelihood ratios for the expected tree that were not significantly lower than those of the best tree obtained in each case (Table 6). This suggests that the expected tree (Figure 4) cannot be statistically ruled out for the tRNA and rRNA separate data sets. This conclusion is also confirmed by the higher bootstrap values (with MP and NJ) obtained for the combined data set of all mitochondrial genes compared with the bootstrap values (MP and NJ) obtained for the separate data set of all protein-coding genes (Figure 4). This stronger support of the expected topology is due to the addition of the phylogenetic information contained in the rRNA and tRNA data sets to that contained in the protein-coding data set.

Phylogenetic position of bichirs: Analyses of the combined protein-coding genes by MP, NJ, and ML

(Figure 4) rule out the hypothesis that the bichir is a sarcopterygian (Figure 1a). Without the addition of an appropriate, and yet unavailable, outgroup, it remains, however, difficult to distinguish between the two hypotheses of bichir as a third major bony fish lineage (Figure 1b) or the birchir as a basal actinopterygian (Figure 1c). One problem in assessing polypterine relationships is that the bichir is one of the most basal vertebrate taxa for which the complete mitochondrial genome is currently available. Lamprey has been shown to be unable to resolve vertebrate relationships and is too distant to be used as outgroup (RUSSO et al. 1996; ZARDOYA and MEYER 1996a,c). Therefore, the phylogenetic position of the bichir can only be determined in the future if members of the Chondrichthyes (sharks, rays and chimaeras) are used as outgroup and a sturgeon mitochondrial genome has been sequenced.

The *Cytochrome b* gene has been sequenced for both the white shark (*Carcharodon carcharias*) and the tiger shark (*Galeocerdo cuvier*) (MARTIN and PALUMBI 1993). Additionally, the sturgeon *cyt b* sequence is also available (BROWN et al. 1989). To further test polypterine relationships, the same analyses as described above were performed with this gene using the same 10 taxa as above plus sturgeon and the two sharks. This data set was analyzed with MP, NJ and ML using both nucleotide sequences (third positions
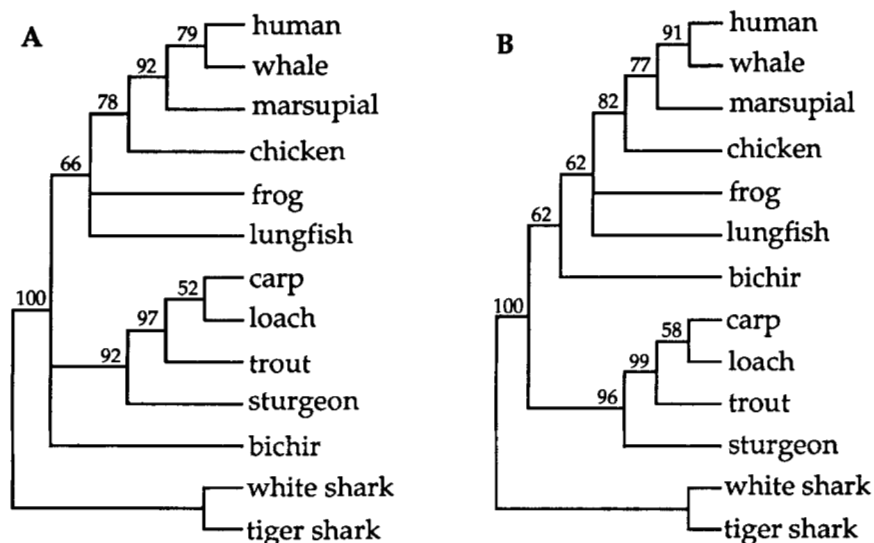
FIGURE 5.—Majority rule bootstrap (FELSENSTEIN 1985) consensus trees of the *cytochrome b* data set. (A) Tree was obtained from MP analysis. (B) Tree was obtained from NJ analysis. Numbers above branches indicate bootstrap values (1000 replicates). ML analysis arrived at the same topology shown for tree A.
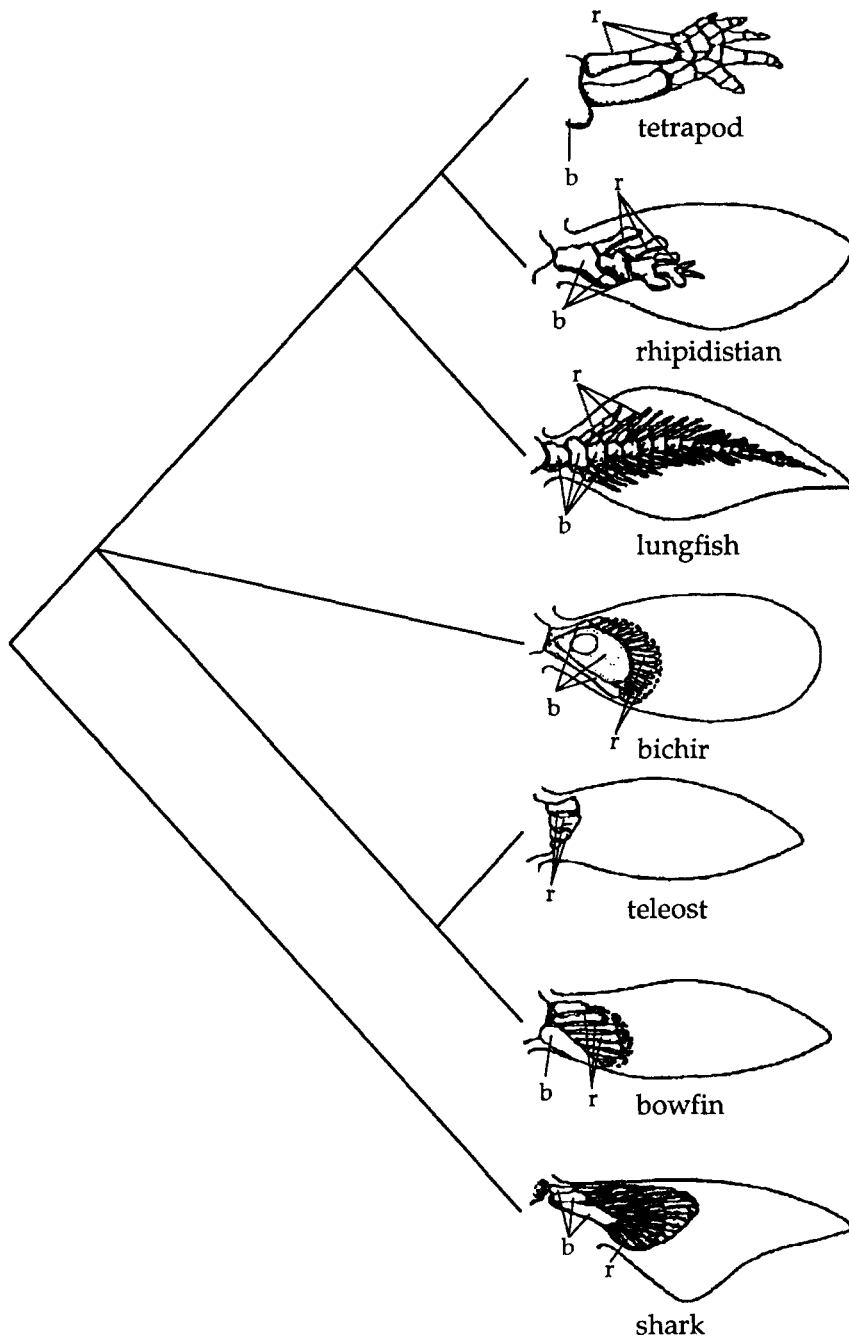
FIGURE 6.—Phylogenetic hypothesis and pectoral limb evolution. The basal (b) and radial (r) bones are indicated in each of the fin/limb schematics. Only up to three basals and radials are indicated, often their number is larger. The homology of some basal bones, *e.g.*, in the birchir and the bowfin (*Amia calva*) is disputed. The homology, and numbers of some of the axial elements between ray-finned and lobe-finned fish is ambiguous (from ROMER and PARSONS 1977).

were excluded) and amino acid sequences and the sharks were declared outgroups. MP and ML analyses place the bichir as sistergroup to all other ray-finned fish included in the analysis; *i.e.*, the bichir is placed basal to sturgeon, supporting the view that bichir should not be placed together with the Acipenseriformes (Table 1) into the Chondrostei (*contra* NELSON 1994). However, the bootstrap value for the node combining bichir with the other ray-finned fish in MP analysis is low (36%), leading to an unresolved trichotomy between bichir, ray-finned fish and lobe-finned fish in a 50% majority rule consensus tree (Figure 5a).

Interestingly, the NJ analysis places the bichir as the sistergroup to lungfish and tetrapods supporting the

theory that the bichir might be a distinct piscine lineage (Brachiopterygii, *sensu* BJERRING 1985) that is more closely related to sarcopterygians than to actinopterygians (Figure 5b). This tentative topology is only weakly supported by a low bootstrap value of 62%.

For all three phylogenetic methods (MP, NJ and ML), the relationships between lungfish and frog are not convincingly resolved with the *cytochrome b* data set. The discordant results and the low bootstrap values for generally well-established tetrapod relationships support the notion that single genes often are unable to obtain the correct phylogeny among such lineages that diverged >400 mya (CAO *et al.* 1994; RUSSO *et al.* 1996; ZARDOYA and MEYER 1996a,c).

Our results show that the bichir is a reliable outgroup to assess relationships among jawed vertebrates, leading to well-supported topologies when large data sets (all mitochondrial genes combined or at least all protein coding genes combined) are assayed. It can be expected that this newly obtained mitochondrial sequence will be especially useful to polarize mitochondrial data sets with which one can hope to investigate the relationships among the most species-rich class of vertebrates. The firm establishment of the phylogenetic position of the bichir awaits the collection of additional data sets, *e.g.,* the complete mitochondrial genomes of an acipenseriform fish (*e.g.,* sturgeon) and a cartilaginous fish.

The resolution of the bichir's phylogenetic position has bearing on the understanding of the evolutionary history, establishment of homology, and the evolution of developmental mechanisms that led from fish fins to vertebrate limbs. Pentadactyl tetrapod limbs have evolved from sarcopterygian lobed fins (*e.g.,* AHLBERG and MILLNER 1994; COATES 1994). The bichir's phylogenetic position would suggest that its lobed fins might not be a shared derived trait that unites them with sarcopterygians but rather that the birchir's lobed fins had evolved independently (Figure 6). Fossil evidence (*e.g.,* reviewed in CARROLL 1988) already indicated that fossil actinopterygian fish had endochondrial radial bones that did not extend far out into the fin. The birchir's probable phylogenetic positions as an actinopterygian would therefore argue for independent evolution of its lobed-fins from the plesiomorphic actinopterygian condition. The disputed homologies (reviewed in *e.g.,* PATTERSON 1980, 1982; AHLBERG and MILLNER 1994; COATES 1994) for the axial elements of the pectoral appendages of fish and land vertebrates and the evolution of their development (Figure 6) will be facilitated by the firm establishment of the phylogenetic position of the bichir.

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.

AHLBERG, P. E., and A. R. MILNER, 1994   The origin and early diversification of tetrapods. Nature **368:** 507–514.

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.

ANDREWS, S. M., 1967   Class Crossopterygii, pp. 641–642 in *The Fossil Record,* edited by W. B. HARLAND. Geological Society, London.

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.

BJERRING, H. C., 1985   Facts and thoughts on piscine phylogeny, pp. 31–57 in *Evolutionary Biology of Primitive Fishes,* edited by R. E. Foreman, A. GORBMAN, J. M. DODD and R. OLSSON. Plenum Press, New York.

BROWN, J. R., T. L. GILBERT, D. J. KOWBELL, P. J. O'HARA, N. E. BUROKER *et al.,* 1989   Nucleotide sequence of the apocytochrome b gene in white sturgeon mitochondrial DNA. Nucleic Acids Res. **17:** 4389.

CAO, Y., J. ADACHI, A. JANKE, S. PAABO and M. HASEGAWA, 1994   Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. J. Mol. Evol. **39:** 519–527.

CARROLL, R. L., 1988   *Vertebrate Paleontology and Evolution.* 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.

COATES, M. I., 1994   The origin of vertebrate limbs. Development (Suppl.) 169–180.

COPE, E. D., 1871   Contribution to the ichthyology of the Lesser Antilles. Trans. Am. Phil. Soc. **14:** 445–483.

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.

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.

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.

GARDINER, B. G., 1967   Further notes on palaeoniscoid fishes with a classification of the Chondrostei. Bull. Br. Mus. (Nat. Hist.) Geol. **14:** 143–206.

GARDINER, B. G., 1973   Interrelationships of teleostomes, pp. 105–135 in *Interrelationships of Fishes,* edited by P. H. GREENWOOD, R. S. MILES and C. PATTERSON. Academic Press, London.

GARDINER, B. G., and A. B. SCHAEFFER, 1989   Interrelationships of lower actinopterygian fishes. Zool. J. Linn. Soc. **97:** 135–187.

GOODRICH, E. S., 1928   Polypterus a palaeoniscid? Palaeobiologica **1:** 87–92.

GREENWOOD, P. H., 1974   Review of Cenozoic freshwater fish faunas in Africa. Ann. Geol. Surv. Egypt **4:** 211–232.

HIXSON, J. E., T. W. WONG and D. A. CLAYTON, 1986   Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand replication. J. Biol. Chem. **261:** 2384–2390.

HUXLEY, T. H., 1861   Preliminary essay upon the systematic arrangement of the fishes of the Devonian epoch. Mem. Geol. Surv. U.K. Decade **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.

JARVIK, E., 1980   *Basic Structure and Evolution of Vertebrates.,* Vol. 1. Academic Press, London.

JESSEN, H. L., 1973   Interrelationships of actinopterygians and brachiopterygians: evidence from pectoral anatomy, pp. 227–232 in *Interrelationships of Fishes,* edited by P. H. GREENWOOD, R. S. MILES and C. PATTERSON. Academic Press, London.

KIMURA, M., 1980   A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

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.

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

LAUDER, G. V., and K. F. LIEM, 1983   The evolution and interrelation-

ships of the actinopterygian fishes. Bull. Mus. Comp. Zool. **150:** 95–197.

Lê, H. L. V., G. Lecointre and R. Perasso, 1993 A 28S based phylogeny of the Gnathostomes: first steps in the analysis of conflict and congruence with morphologically based clado-grams. Mol. Phyl. Evol. **2:** 31–51.

Lee, W.-J., W. Conroy, H. Howell and T. D. Kocher, 1995 Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. **41:** 54–66.

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.

Martin, A. P., and S. R. Palumbi, 1993 Protein evolution in different cellular environments: cytochrome b in sharks and mammals. Mol. Biol. Evol. **10:** 873–891.

Meyer, A., 1993 Evolution of mitochondrial DNA in fishes, pp. 1–38 in *Biochemistry and Molecular Biology of Fishes*, edited by H. A. 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.

Naylor, G. J., T. M. Collins and W. M. Brown, 1995 Hydrophobicity and phylogeny. Nature **373:** 555–556.

Nelson, G. J., 1973 Relationships of clupeomorphs, with remarks on the structure of the lower jaw in fishes, pp. 333–349 in *Interrelationships of Fishes*, edited by P. H. Greenwood, R. S. Miles and C. Patterson. Academic Press, London.

Nelson, G. J., 1994 *Fishes of the World.* John Wiley & Sons, New York.

Normark, B. B., A. R. McCune and R. G. Harrison, 1991 Phylogenetic relationships of neopterygian fishes, inferred from mitochondrial DNA sequences. Mol. Biol. Evol. **8:** 819–834.

Ortí, G., and A. Meyer, 1996 Molecular evolution of Ependymin and the phylogenetic resolution of early divergences among euteleost fishes. Mol. Biol. Evol. **13:** 556–573.

Patterson, C., 1980 Origin of tetrapods: historical introduction to the problem, pp. 159–175 in *The Terrestrial Environment and the Origin of Land Vertebrates*, edited by A.L. Panchen. Academic Press, London.

Patterson, C., 1982 Morphology and interrrelationships of primitive actinopterygian fishes. Am. Zool. **22:** 241–259.

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., and T. S. Parsons, 1977 *The Vertebrate Body*, Ed. 5, Saunders Company, Philadelphia.

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.

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.

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

Schaeffer, B., 1973 Interrelationships of chondrosteans, pp. 207–226 in *Interrelationships of Fishes*, edited by P. H. Greenwood, R. S. Miles and C. Patterson. Academic Press, London.

Shubin, N. H., and P. Alberch, 1986 A morphogenetic approach to the origin and basic organization of the tetrapod limb, pp. 319–387 in *Evolutionary Biology*, edited by M. K. Hecht, B. Wallace, and G. I. Prance. Plenum Press, New York.

Sordino, P., and D. Duboule, 1996 A molecular approach to the evolution of vertebrate paired appendages. Trends Ecol. Evol. **11:** 114–119.

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.

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

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.

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.

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.

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 The origin of tetrapods based on 28S ribosomal RNA sequences. 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.