Molecular Evolution of Ependymin and the Phylogenetic Resolution of Early Divergences Among Euteleost Fishes

Guillermo Ortí[1] *and Axel Meyer*

Department of Ecology and Evolution, State University of New York at Stony Brook

The rate and pattern of DNA evolution of ependymin, a single-copy gene coding for a highly expressed glycoprotein in the brain matrix of teleost fishes, is characterized and its phylogenetic utility for fish systematics is assessed. DNA sequences were determined from catfish, electric fish, and characiforms and compared with published ependymin sequences from cyprinids, salmon, pike, and herring. Among these groups, ependymin amino acid sequences were highly divergent (up to 60% sequence difference), but had surprisingly similar hydropathy profiles and invariant glycosylation sites, suggesting that functional properties of the proteins are conserved. Comparison of base composition at third codon positions and introns revealed AT-rich introns and GC-rich third codon positions, suggesting that the biased codon usage observed might not be due to mutational bias. Phylogenetic information content of third codon positions was surprisingly high and sufficient to recover the most basal nodes of the tree, in spite of the observation that pairwise distances (at third codon positions) were well above the presumed saturation level. This finding can be explained by the high proportion of phylogenetically informative nonsynonymous changes at third codon positions among these highly divergent proteins.

Ependymin DNA sequences have established the first molecular evidence for the monophyly of a group containing salmonids and esociforms. In addition, ependymin suggests a sister group relationship of electric fish (Gymnotiformes) and Characiformes, constituting a significant departure from currently accepted classifications. However, relationships among characiform lineages were not completely resolved by ependymin sequences in spite of seemingly appropriate levels of variation among taxa and considerably low levels of homoplasy in the data (consistency index = 0.7). If the diversification of Characiformes took place in an "explosive" manner, over a relatively short period of time this pattern should also be observed using other phylogenetic markers. Poor conservation of ependymin's primary structure hinders the design of efficient primers for PCR that could be used in wide-ranging fish systematic studies. However, alternative methods like PCR amplification from cDNA used here should provide promising comparative sequence data for the resolution of phylogenetic relationships among other basal lineages of teleost fishes.

Introduction

Of the vast array of genes available for molecular phylogenetic studies, only a small subset has been investigated so far. By and large, mitochondrial DNA (mtDNA) and a handful of nuclear genes, typically ribosomal RNA genes, have been used for phylogenetic reconstruction. Most research on animals that used mitochondrial DNA sequences has dealt with populations or low taxonomic level relationships (reviewed in Wilson et al. 1985; Moritz, Dowling, and Brown 1987; Avise et al. 1988; Meyer 1994*a*; but see Meyer and Wilson 1990). Ancient divergences (i.e., older than 250 million years ago) have been mostly addressed using nuclear ribosomal gene sequences (e.g., Woese and Fox 1977; Woese, Kandler, and Wheelis 1990; Field et al. 1988; Sogin, Edman, and Elwood 1989; Hedges, Moberg, and Maxson 1990). Doubtless, ease of data collection and a wealth of comparative data for these two classes of frequently used genetic markers have contributed to their continued use as phylogenetic tools of choice. However, as empirical results have accumulated, limitations of these phylogenetic markers have become apparent (e.g., Smith 1989; Normark, McCune, and Harrison 1991; Stock et al. 1991). Lately, phylogenies based on protein-coding genes for the study of early divergences have been shown to provide robust estimations (e.g., Hashimoto et al. 1994; Sidow and Thomas 1994). Only recently has the phylogenetic information content of an increasing diversity of protein-coding nuclear genes been investigated. These excursions into new "gene space" by molecular systematists searching for alternative tools to assess intermediate and higher level taxonomic relationships (Friedlander, Regier, and Mitter 1992, 1994; Graybeal 1994; Soto-Adames, Robertson, and Berlocher 1994) have produced some promising results. We here explore the phylogenetic utility and molecular evolution of ependymin, a single-copy nuclear gene, and its application in high order systematics of characiforms and other otophysan fishes (electric fishes, catfishes, and carps, fig. 1) thought to have originated about 140 million years ago (Lundberg 1993).

Ependymin is a secretory product of meningeal cells and forms a major glycoprotein component of the cerebrospinal fluid in various orders of teleost fish. Originally identified in the goldfish brain (in the ependymal zone) in studies of the role of extracellular proteins in neuroplasticity and learning (Shashoua 1976; reviewed in Shashoua 1985, 1988, 1991; Hoffmann 1994), its genetic basis has been determined in salmoniform, clupeiform and cypriniform fishes (Adams and Shashoua 1994; Müller-Schmid et al. 1992, 1993; Rinder et al. 1992). The ependymin protein is encoded by a single-copy gene, but two copies (presumably representing duplicated loci) have been found in the gold-

[1] Present address: Department of Genetics, University of Georgia.

Key words: Otophysi, Characiformes, Protacanthopterygii, ependymin, cDNA sequences, RT-PCR, codon bias, molecular phylogeny, biogeography.

Address for correspondence and reprints: Guillermo Ortí, Department of Genetics, University of Georgia, Athens, Georgia 30602. E-mail: gorti@bscr.uga.ed.

Mol. Biol. Evol. 13(4):556–573. 1996

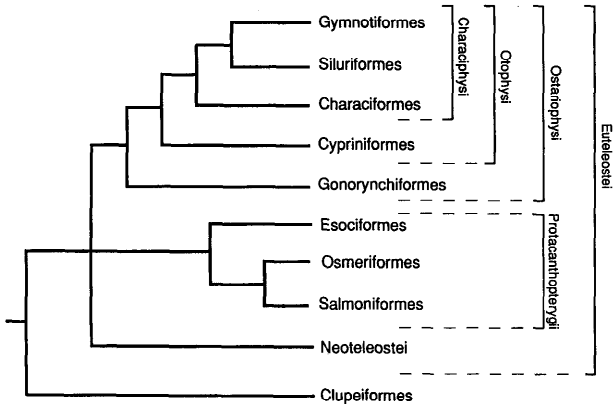

FIG. 1.—Taxonomic arrangement and phylogenetic relationships of the superorders Ostariophysi and Protacanthopterygii, according to Nelson (1994) and Fink and Fink (1981). The latter authors, however, placed siluriforms and gymnotiforms as suborders of the order Siluriformes.

fish (Königstorfer et al. 1989; Königstorfer, Sterrer, and Hoffmann 1989) and rainbow trout (Müller-Schmid et al. 1993), which belong to tetraploid families. The encoded precursor of ependymin is 212–221 amino acids long, organized into six exons (fig. 2), deduced by the comparison of cDNA and genomic sequences in zebrafish, carp, goldfish, and Atlantic salmon (Rinder et al. 1992; Müller-Schmid et al. 1992; Adams and Shashoua 1994). Ependymin has been shown to have a high turnover rate in the brain after experimental fish were subjected to classical conditioning experiments, and was therefore thought to be involved in long-term memory consolidation (e.g., Shashoua 1991). Increased expression of ependymin has also been reported during optic-nerve regeneration (Schmidt and Shashoua 1988). Although the molecular function of ependymin is not fully understood, it shares features with soluble glycoproteins that are involved in cell contact phenomena (Hoffmann 1992). It is usually present as a disulfide-linked dimeric protein with several glycosylation variants differing in carbohydrate content, and has the capacity to bind Ca^{2+}, presumably leading to functionally significant conformational transitions (Shashoua 1991; Hoffmann 1994).

Otophysan fishes comprise a well-known lineage of teleosts including cypriniforms (carps, loaches, minnows, suckers), siluriforms (catfishes), gymnotiforms (South American knife fishes or electric fishes), and characiforms (African and South American tetras, piranhas, hatchet fishes, headstanders, etc.). They are defined by specialized modifications in their anteriormost four or five vertebrae, and by the possession of movable bony elements connecting the inner ear with the swim bladder (the Weberian apparatus). Phylogenetic relationships among them have so far been addressed using morphological traits only. The currently accepted hypothesis of relationships among the orders of otophysan fish was advanced by Fink and Fink (1981, fig. 1). Few explicit phylogenetic studies have been published proposing relationships among all the major lineages within each order (e.g., Alves-Gomes et al. 1995), in spite of the great importance attached to these fishes for biogeographic studies (Myers 1938, 1949; Lundberg 1993). Characiforms are primary freshwater fishes found only in Africa and South America, which are thought to have separated by the opening of the South Atlantic ocean some 84–106 million years ago (Parrish 1993; Pitman et al. 1993). A major question arising from their present biogeographic distribution is whether all Neotropical characiforms and all African characiforms form separate monophyletic groups as a consequence of this massive vicariant event. While the African characiform fauna consists of approximately 200 species assigned to four families, characiforms are much more diverse in South America, where the estimated number of species is over 1,100, placed in 12 nominal Neotropical families (Greenwood et al. 1966). A preliminary phylogenetic study based on morphology (Buckup 1991) suggested that at least three sister-group relationships exist between African and Neotropical characiforms, implying either unlikely marine dispersal events or a disproportionate extinction rate of African characiforms following the continental split (Lundberg 1993). We will use ependymin sequence data to examine the relationships between African and Neotropical characiform lineages in relation to these intriguing biogeographic scenarios.

At higher taxonomic levels, interrelationships of otophysans, protacanthopterygians, and neoteleostean fishes are still largely unsettled (Rosen and Greenwood 1970; Rosen 1974, 1985; Fink and Weitzman 1982; Lauder and Liem 1983; Fink 1984; Begle 1991, 1992; Nelson 1994). In spite of much effort to classify these fishes, interpretation of the morphological evidence has been difficult at best, and much disagreement persists. The taxonomic arrangement adopted by Nelson (1994) is shown in figure 1. We here attempt to establish a molecular phylogeny for otophysans, and characiform lineages in particular, while addressing the utility of ependymin as a genetic marker for higher level teleostean systematics. This study is based on newly determined DNA sequences of ependymin for selected species of electric fishes, catfishes, and characiforms, and on published DNA sequences for cyprinids (carp, goldfish, and zebrafish), salmonids, pike and herring (table 1).

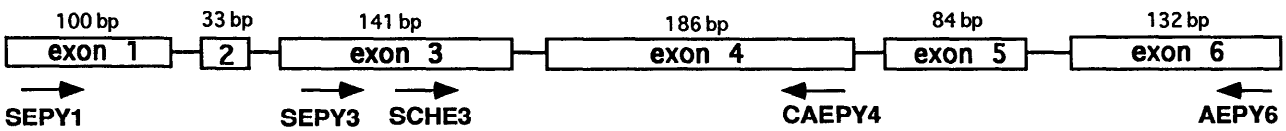

FIG. 2.—Ependymin gene structure and location of PCR primers used for this study (for primer sequences see text). Intron-exon organization (and exon lengths) are based on the zebrafish gene.

Table 1
List of Sequences and Systematic Position of Fish Used in This Study

Organism (Order, Family, Genus)		GenBank Accession Number
Order Characiformes		
Family Alestidae:	*Alestes*	U33475[a]
	Phenacogrammus	U33476[a]
Family Distichodontidae:	*Distichodus*	U33477[a]
Family "Characidae":	*Chalceus*	U33478[a]
	Paracheirodon	U33479[a]
	Gymnocorymbus	U33480[a]
	Metynnis	U33481[a]
Family Gasteropelecidae:	*Gasteropelecus*	U33482[a]
Family Anostomidae:	*Leporinus*	U33483[a]
Family Hemiodontidae:	*Hemiodus*	U33484[a]
Family Erythrinidae:	*Hoplias*	U33485[a]
Family Ctenoluciidae:	*Boulengerella*	U33486[a]
Family Lebiasinidae:	*Nannobrycon*	U33487[a]
Order Siluriformes		
Family Loricariidae:	*Hypostomus*	U33488[a]
Family Pimelodidae:	*Pimelodus*	U33489[a]
Family Schilbeidae:	*Schilbe*	U33490[a]
Family Mochokidae:	*Synodontis*	U33491[a]
Order Gymnotiformes		
Family Eigenmanniidae[b]:	*Eigenmania*	U33492[a]
Family Rhamphichthyidae:	*Rhamphichthys*	U33493[a]
Order Cypriniformes		
Family Cyprinidae:	*Cyprinus carpio*	U00432
	Carassius auratus	U00433, X14134
	Danio rerio	M89643
Order Salmoniformes		
Family Salmonidae:	*Salmo salar*	M93699
Order Esociformes		
Family Esocidae:	*Esox lucius*	L09066
Order Clupeiformes		
Family Clupeidae:	*Clupea harengus*	L09065

[a] Determined for this study.
[b] Following Alves-Gomes et al. (1995).

Materials and Methods

Laboratory Procedures

Live fish were obtained from commercial sources (table 1) and stored at −80°C. Brains were dissected out of frozen fish, kept on dry ice and subsequently stored at −80°C. Translationally active RNA was extracted from the brain tissue of one or more individuals per species by homogenization with a Polytron blender (Janke and Kunkel AG Bresiga, Germany) in 1 mL of lysis buffer containing 5 M guanidine thiocyanate, following the protocol of Cathala et al. (1983). Briefly, after homogenization, 7 mL of 4 M LiCl solution was added for overnight incubation at 4°C. Large DNA molecules were sheared by pumping the volume several times through an 18½ gauge needle attached to a sterile syringe. RNAs were then precipitated by centrifugation at 11,000 × *g* at 4°C and the pellet was resuspended in 0.5 mL of buffer containing 1 mM EDTA and 0.2% SDS, and extracted with phenol : chloroform. The purified product was obtained by ethanol precipitation and resuspension in depc-H_20, and its concentration was measured with a spectrophotometer.

Reverse transcription of 2 μg of the purified RNA to cDNA was performed in 40 μL reactions with MuLV reverse transcriptase using a GeneAmp RNA PCR kit (Perkin Elmer) following the manufacturer's protocol. To increase specificity in the subsequent PCR reactions only mRNAs with poly rA tails were transcribed by using Oligo $d(T)_{16}$ primer in the cDNA synthesis. A 1:5 dilution of the cDNA product so obtained was used as template for PCR amplifications (Saiki et al. 1985).

Primers for PCR (Saiki et al. 1985; Mullis et al. 1986) were designed based on published zebrafish and goldfish ependymin sequences (fig. 2) (Königstorfer et al. 1989; Königstorfer, Sterrer, and Hoffmann 1989; Sterrer, Königstorfer, and Hoffmann 1990). All samples were first PCR-amplified with the primer pair SEPY1-AEPY6, located in exons 1 and 6 respectively ("sense" primer SEPY1: 5′-GTCAAGCTGCTCTGTGTGG, and "antisense" primer AEPY6: 5′-TCGTGGAACA-GGTCGAAGAA). Double-stranded PCR amplifications were performed in 30-μL volumes containing 67 mM Tris (pH 8.8), 6.7 mM $MgCl_2$, 16.6 mM $(NH_4)_2SO_4$, 10 mM 2-mercaptoethanol, 0.5 mM of each dNTP, 1 μM of each primer, 0.5 units of *Taq* polymerase (Perkin Elmer), and 5 μL of the cDNA 1:5 dilution. Amplification products (around 630 bp long, variable among species) were gel-purified (2.5% Nusieve-Agarose in TAE buffer), and used as template for the generation of single-stranded DNA for direct sequencing by asymmetric PCR (Gyllensten and Erlich 1988). Single-stranded DNA was

concentrated and desalted in spin columns (Millipore: Ultrafree-MC30,000) and sequenced from both directions with the same PCR primers by the dideoxy method using a commercial kit (Sequenase, United States Biochemical). In those cases where the SEPY1-AEPY6 primer pair did not yield a single PCR product (or the sequencing reactions did not produce readable sequences in excess of 300 bp) a second double-stranded amplification was attempted using the gel-purified products and internal primers SCHE3 and CAEPY4 in the following combination: SEPY1-CAEPY4 and SCHE3-AEPY6 (CAEPY4: 5′-GGACTTGGACGGGCTGCC, located in exon 4, and SCHE3: 5′-GAGAAAGAAAACCACTCAAACAAAAC, located in exon 3). Clean single products obtained were used as template for asymmetric PCR and direct sequencing as above. In some cases (for *Eigenmania, Hypostomus, Hoplias,* and *Paracheirodon*) where these procedures failed to yield clean PCR products, the primer SEPY3 was used instead of SEPY1 for double-stranded amplifications with AEPY6. SEPY3 is located in a conserved region of the third exon (SEPY3: 5′-GGNGARTTYAGNTAYGAYTC, where N = G, A, T, or C; Y = T or C; and R = A or G). A single double-stranded product 500 bp long obtained was gel-purified and directly sequenced from both directions, as above.

Data Analysis

DNA sequences obtained were aligned by eye with the multiple-sequence editor ESEE (Cabot and Beckenbach 1989), and translated to amino acid sequences using the universal genetic code. Pairwise comparisons and statistical information from the sequences were obtained with MEGA version 1.0 (Kumar, Tamura, and Nei 1993), as well as phylogenetic trees based on the neighbor-joining (NJ) method (Saitou and Nei 1987). Phylogenetic inference by maximum parsimony analyses were performed using PAUP version 3.1.1 (Swofford 1993) and MACCLADE version 3 (Maddison and Maddison 1992); maximum likelihood analyses (Felsenstein 1981) used FASTDNAML version 1.0.8 (Olsen et al. 1994) and MOLPHY version 2.2 (Adachi and Hasegawa 1994). Specific weighting schemes adopted for different codon positions and kinds of substitution were derived from preliminary analyses of the substitution patterns (based on pairwise contrasts) and are described in each case where results are presented. Bootstrapping (Felsenstein 1985) was used to estimate confidence in the results. Previously published ependymin sequences were retrieved from GenBank for Cypriniformes (*Carassius, Danio,* and *Cyprinus*; Königstorfer et al. 1989; Königstorfer, Sterrer, and Hoffmann 1989; Sterrer, Königstorfer, and Hoffmann 1990; Adams and Shashoua 1994) and included in the analyses. Protacanthopterygians (*Salmo* and *Esox*) and herring (*Clupea*) were used as outgroups (Müller-Schmid et al. 1993; table 1). Hydropathy profiles were computed with the algorithm of Kyte and Doolittle (1982), using the program MACVECTOR version 4.1.1 (Eastman Kodak).

Results and Discussion

Variation Among Ependymin Sequences

A fragment of about 600 bp of cDNA sequence was obtained for most taxa listed in table 1. Sequences from the 5′ end (150 bp) could not be determined for six specimens (fig. 3) due to low versatility of primer SEPY1. The first 21 amino acid residues at the N-terminus of the ependymin precursor contain the signal peptide sequence (where the SEPY1 primer is located, fig. 2) which is not entirely conserved across taxa (Müller-Schmid et al. 1993). Amino acid sequences inferred from the cDNA sequences were aligned with published sequences from cyprinids, salmoniforms, and the herring (fig. 3). Percent sequence differences among amino acid and DNA sequences were large (without correction for multiple hits, table 2), confirming previous observations that ependymin is a rapidly evolving gene (Müller-Schmid et al. 1993). Even cysteine residues are not fully conserved, but the pattern of variation seems to agree with major taxonomic divisions. For example, cysteine residues are found at position 20 for catfish and at positions 154–155 for most (but not all) catfish, gymnotids, and characiforms (fig. 3). Length variation among sequences is also present, and is most noteworthy in comparisons involving the catfish (fig. 3). The 4–8 amino acid residues accounting for length variation are located in the proximity of intron 4, deduced from the conserved intron-exon structure of ependymin in goldfish, carp, zebrafish and salmon (Müller-Schmid et al. 1992; Rinder et al. 1992; Adams and Shashoua 1994). *Distichodus* and *Nannobrycon* share a deletion at position 48 (fig. 3). Conserved features in the sequences including potential N-glycosylation sites and cysteine and tryptophan residues are also shown in figure 3. The most conserved region is located around the potential N-glycosylation site at position 80 (figs. 3 and 4). This site is presumably necessary for binding crucial oligosaccharide units and conferring calcium-binding capacity to the molecule (Müller-Schmid et al. 1993). Notice the additional potential N-glycosylation site at position 139 shared by characiform sequences (except *Distichodus*), which is not found in any other species.

Base Composition and Nucleotide Substitution Pattern

As expected, at the nucleotide level, substitutions at third codon positions accumulate faster than at first and second positions (table 2, fig. 5). Low levels of change at second positions may be related to the observed strong conservation of hydrophobicity profiles (fig. 4) and to base composition (fig. 6). In spite of their low levels of similarities (up to 60% sequence difference in amino acid composition, table 2) ependymin amino acid sequences show surprisingly similar hydropathy profiles (fig. 4), suggesting that functional properties are well conserved among these highly divergent proteins. As suggested by Naylor, Collins, and Brown (1995), a strong functional requirement for hydrophilicity effectively constrains character-state space in second positions of codons. Base composition at second positions

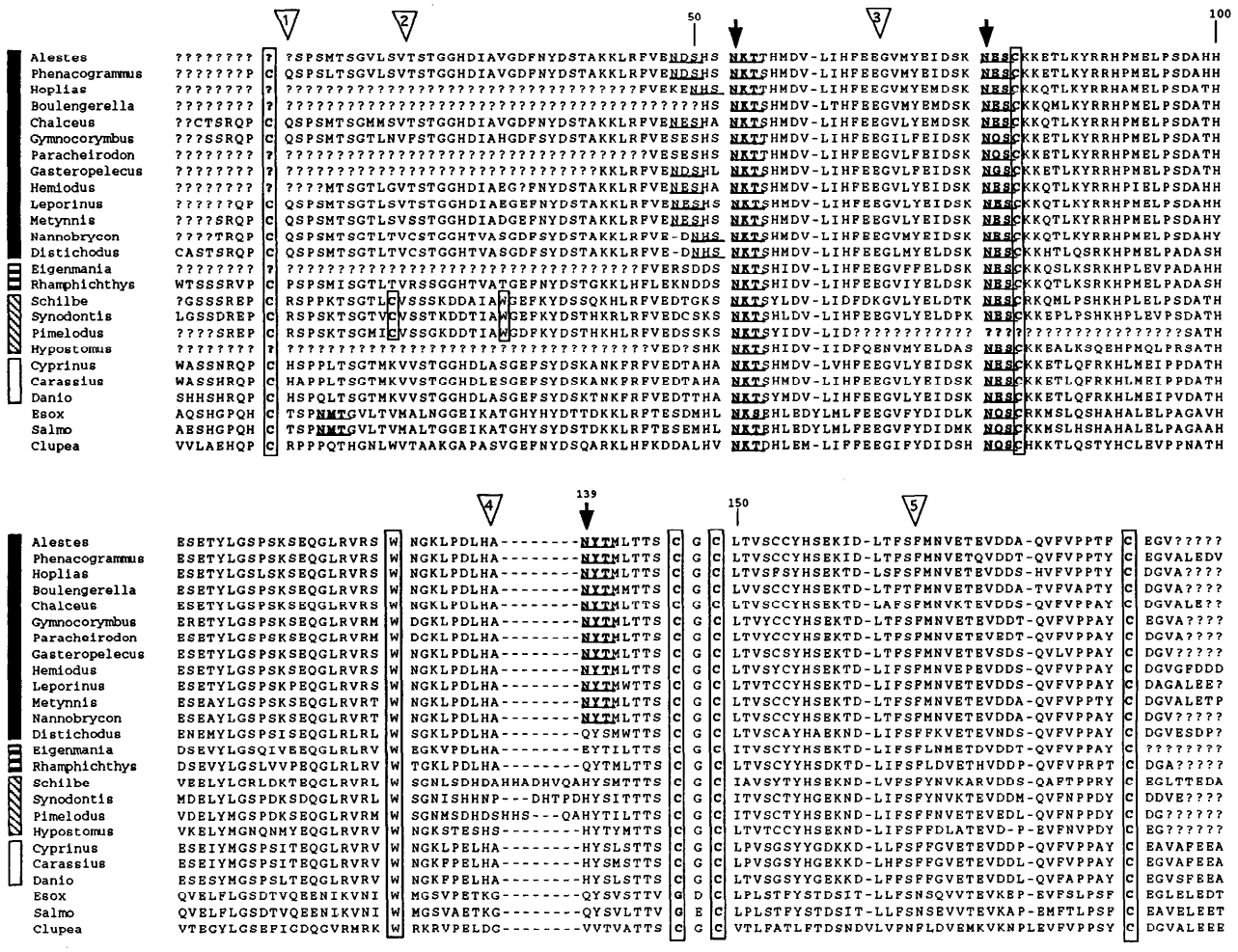

FIG. 3.—Alignment of partial ependymin protein sequences used in this study. Conserved features are marked: tryptophan and cysteine residues are boxed, potential glycosylation sites (NXT and NXS) are marked by arrows and underlinecd. Notc glycosylation site at position 139 common to all characiforms except *Distichodus* and *Nannobrycon,* and at position 13 common to salmon and pike. Inverted triangles indicate position of introns, based on the zebrafish gene structure. A black bar next to the taxon names marks all characiform taxa, whereas bars with horizontal lines, cross-hatched, and empty indicate electric-fish, catfish, and cyprinid taxa, respectively.

is strongly A-biased (fig. 6): triplets with an A in the second position are relatively more frequent (33.5% mean frequency, table 3) than the other classes. All triplets with an A in the second position code for hydrophilic amino acids (table 3; Kyte and Doolittle 1982), which is expected for a soluble protein. In mitochondrial genes that code for protein subunits with abundant membrane-spanning domains, a high proportion of triplets have either C or T in second positions encoding for hydrophobic residues (Naylor, Collins, and Brown 1995).

At third codon positions, observed values of more than 60% uncorrected sequence differences strongly suggest that changes among ependymin sequences are saturated (table 2, fig. 5). Saturation would argue in favor of down-weighting or excluding third positions, particularly in phylogenetic analyses involving the deeper nodes. As a first approximation, the number of phylogenetically informative sites at third codon positions should increase with sequence difference but only up to about 20%–30% divergence (DeSalle et al. 1987; Friedlander, Regier, and Mitter 1994). Higher divergence values would seem to imply a loss of information due to multiple hits in mostly synonymous sites, and therefore suggest that these sites might be phylogenetically uninformative or misleading. However, in their assessment of information content for nuclear genes, Friedlander, Regier, and Mitter (1994) have shown that data sets consisting of third codon positions only, with saturating divergence levels (i.e., above 30%), were still able to recover the "correct" test phylogeny, suggesting that this threshold might be too conservative (see below). In contrast to mitochondrial genes (e.g., Brown and Simpson 1982; Brown et al. 1982), the cpendymin sequences accumulate transversions faster than transitions, except among the least divergent sequences, where most changes are third position transitions (fig. 5). Therefore, as an alternative strategy to down-weighting or eliminating third positions altogether, we also considered only transversions at third positions in some of the parsimony analyses.

Table 2
Ependymin Amino Acid and DNA Percent Sequence Differences within and between Taxa

		Between				
Taxa	Within	Gymnotoidei	Siluroidei	Cyprinidae	Protacantho-pterygii	Clupea
Characiformes	**2.7–23.2**	**23.6–28.7**	**31.9–44.2**	**23.2–30.8**	**52.5–57.1**	**49.7–52.8**
	0.5–15.3	11.7–17.3	13.8–31.6	13.3–21.9	28.6–40.3	28.1–39.8
	0.0–12.2	6.6–14.3	10.2–24.0	8.2–13.3	24.5–35.2	18.9–29.1
	5.6–35.7	26.5–41.8	22.4–46.9	29.1–49.5	41.3–61.2	43.4–60.7
Gymnotoidei	**17.4**		**34.7–43.7**	**31.2–37.2**	**55.6–57.1**	**50.7–52.9**
	11.2		16.3–30.6	17.3–23.5	29.6–40.3	29.1–39.3
	4.6		10.7–23.5	8.7–18.4	21.9–34.7	17.9–29.6
	24.0		27.6–45.4	34.7–53.6	43.9–60.7	46.4–58.2
Siluroidei	**22.7–47.2**			**36.1–46.2**	**59.0–67.5**	**55.6–61.0**
	12.8–28.6			18.4–32.7	30.1–49.5	33.2–46.9
	7.7–18.9			14.8–24.5	23.5–42.3	21.9–33.7
	20.4–33.7			34.7–60.7	43.4–66.8	48.0–63.8
Cyprinidae	**5.6–10.2**				**54.1–58.7**	**51.5–54.6**
	3.1–6.6				37.2–40.3	37.8–39.3
	2.0–4.6				34.2–37.2	26.0–27.0
	12.2–21.9				59.2–64.3	61.2–62.2
Protacanthoptertygii	**10.2**					**58.7**
	4.6					40.8–42.9
	4.1					39.8
	20.4					59.7–61.2

NOTE.—Values from top to bottom are ranges of mean differences for amino acid sequences (bold), first, second, and third codon position nucleotide differences, respectively (uncorrected values).

Synonymous Codon Usage

Also unlike most vertebrate mitochondrial genes (e.g. Tzeng et al. 1992; Meyer 1993), base composition at third codon positions in ependymin sequences have a high GC content and a very low content of A (fig. 6). Most frequently used synonymous codons have either C or G in their third position (table 3). However, cyprinids differ in base composition from the other taxa in that they have a higher T than C or G content in third codon positions (fig. 6). This difference might be due to a different "codon dialect" (Grantham et al. 1980; Ikemura 1985) used by cyprinids or a different overall genomic AT content (Bernardi and Bernardi 1986), or just a consequence of the particular amino acid composition of cyprinid ependymin. Base composition of introns (fig. 6) of salmonid and cyprinid ependymins are likewise AT-rich, but *Salmo* and *Esox* are different from cyprinids at third positions, suggesting that overall genomic base composition may not be responsible for the difference at third positions. Moreover, amino acid content does not differ significantly between cyprinids and the other taxa. Codon usage in cyprinids seems to be different and may account for the observed differences in base composition. Relative synonymous codon usage values are higher than average in cyprinids for F, L, V, S, Y, N, and R codons with a T at the third position. Furthermore, some codons with C or G in third positions are used less frequently than average in cyprinids.

To investigate codon dialect differences further and test whether the observed pattern may represent a more widespread genomic property, we also compared intron and third codon position base compositions for other genes in these taxa (GenBank sequences were obtained for salmon, carp and catfish growth hormone, carp actin and *Astyanax* visual pigments—accession numbers L04688, M94348, S629215, M24113, M90400-3, respectively). In all cases AT-rich intron sequences but GC-rich third codon positions were observed (data not shown), matching the unusual pattern seen in figure 6 for all but cyprinid ependymins. Larger samples will have to be compared in order to determine the significance of the divergent codon usage observed in cyprinid ependymin genes.

Maximum Parsimony and Neighbor-joining Analyses of Nucleotide Sequences

A total of 588 bp were aligned for all 25 taxa, of which 442 were variable and 359 were phylogenetically informative. When third codon positions were excluded only 258 sites were variable and 193 were phylogenetically informative. The herring (*Clupea harengus*) was used as the outgroup. Parsimony analyses were performed with PAUP using heuristic searches (TBR branch-swapping) with 10 random-addition-sequence replications. Bootstrap analyses were based on 100 pseudoreplicates, with heuristic searches with three random-addition-sequence replications. Results differed slightly based on which weighting strategy was adopted (see below). Neighbor-joining analyses were performed with MEGA using Kimura (1981) distances.

Most-parsimonious trees obtained by excluding transitions in third codon positions and by eliminating third positions completely were mostly congruent with each other, but differed somewhat from the ones using all characters with equal weight (fig. 7*A* and *B*). The most basal branches on the trees resulting from all weighting strategies are congruent and receive very high bootstrap support. Protacanthopterygii (*Esox* + *Salmo*), Otophysi (cyprinids + characiforms + siluriforms + gymnotiforms), Characiphysi (characiforms + siluri-

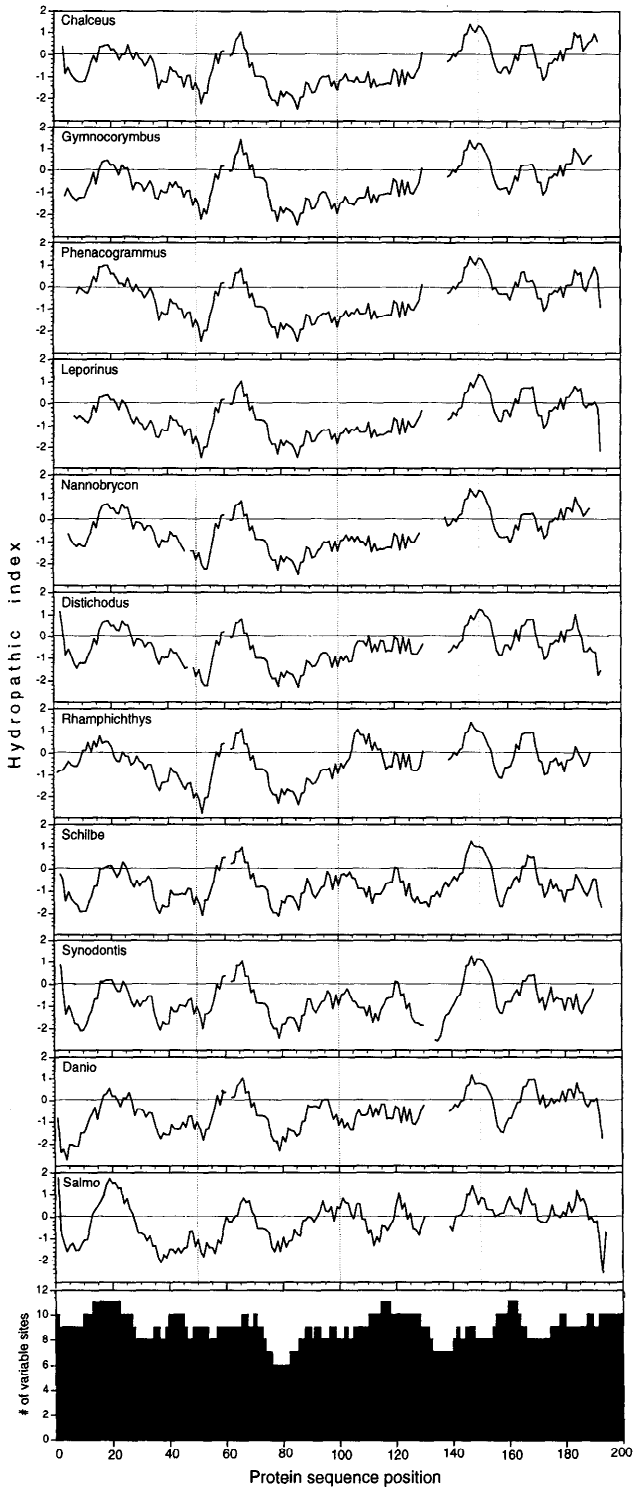

FIG. 4.—Comparison of hydropathic profiles for selected taxa sequenced in this study, plus *Danio* and *Salmo*. Hydropathic index was computed according to Kyte and Doolittle (1982) with a window size of 11 residues. Breaks in hydropathy curves correspond to alignment gaps. Bottom panel shows the number of variable sites in a sliding window of 11 residues.

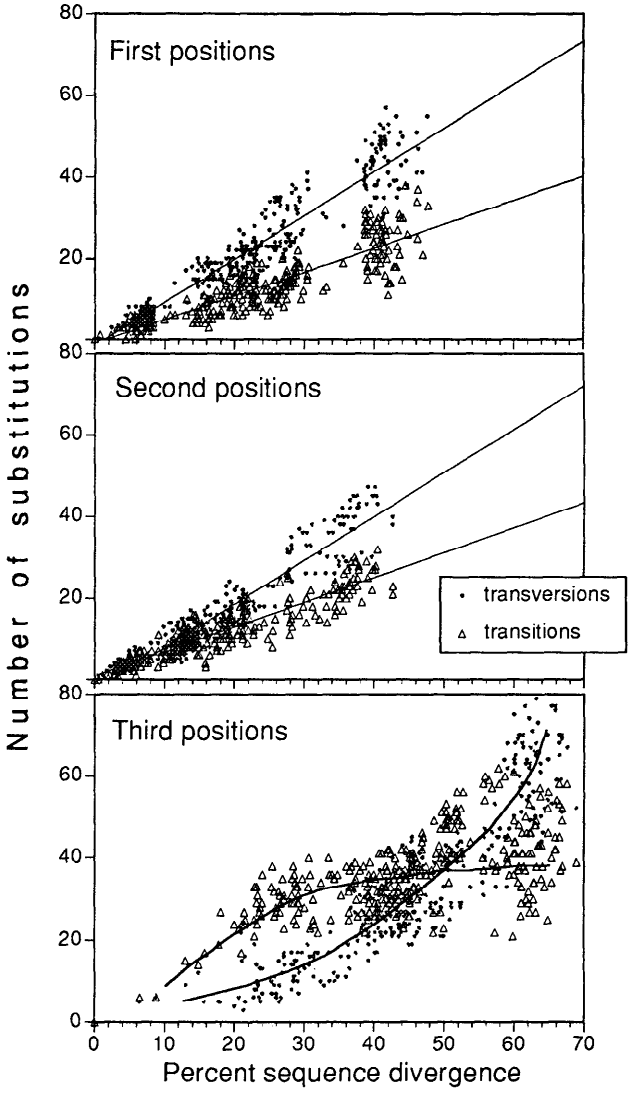

FIG. 5.—Substitution pattern of transitions and transversions for each codon position. The number of transitions and transversions is plotted against total sequence divergence (uncorrected) for all pairwise comparisons of taxa.

forms + gymnotiforms), Gymnotiformes, and Cyprinidae are all strongly supported. In contrast, characiform monophyly is not well supported since the African *Distichodus* tends to branch off before either electric fish (fig. 7*B*) or siluriforms + gymnotiforms (fig. 7*A*). Although the topology within characiforms is congruent for both trees, it receives very low bootstrap support except for the grouping of *Alestes* + *Phenacogrammus* (African family Alestidae) and *Gymnocorymbus* + *Paracheirodon* (Neotropical family Characidae). Note that *Chalceus* and *Metynnis,* traditionally included in the family Characidae, come out in separate branches while *Gasteropelecus* (family Gasteropelecidae) groups with *Gymnocorymbus* + *Paracheirodon.* The most important difference between the trees is the relationship between electric fish, catfish, and characiforms. Whereas tree A (fig. 7) groups catfish with electric fish (the "traditional" hypothesis also shown in fig. 1), tree B (fig. 7) suggests a closer relationship between electric fish and char-

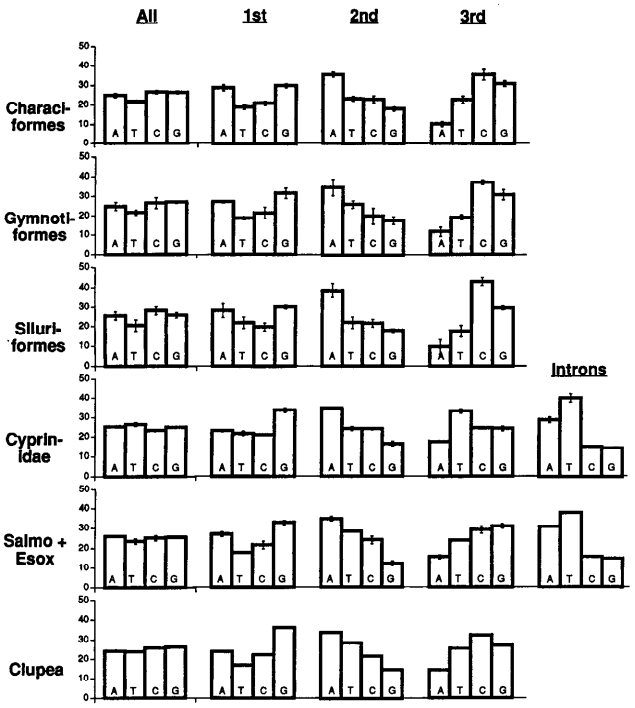

FIG. 6.—Percent base composition by codon position, for each major taxonomic division. Bars are standard deviations. Intron base composition is shown for cyprinids and *Salmo*.

aciforms. Both of these topologies are stable to *a posteriori* reweighting ("successive approximation"; Farris 1969; Carpenter 1988). One of the shortest trees is obtained when this procedure is applied to the data set with all characters equally weighted (fig. 7*A*, with *Distichodus* branching off before Siluriformes + Gymnotiformes). Likewise, reweighting the data set with third codon positions excluded results in a single shortest tree identical to tree B (fig. 7). Of these alternative hypotheses tree A (fig. 7) is less well resolved and has lower bootstrap values and a lower consistency index than tree B (fig. 7), as a likely consequence of considering "noisy" third codon positions (see above and fig. 5). Furthermore, forcing the topology shown in figure 7*B* on the data set with all characters equally weighted required only four additional steps (L = 1,546), in contrast to eight additional steps required by the topology shown in figure 7*A* (L = 1,037) on data excluding transitions in third positions. Excluding the fast evolving third codon positions also results in higher bootstrap support for grouping the electric fish with characiforms (fig. 7*B*) rather than with catfish (fig. 7*A*).

An alternative approach to test for how well particular clades are supported by the data is by inspection of suboptimal trees ("decay analysis or Bremer support," Bremer 1988), counting how many extra steps are required to collapse the clade of interest. For the clade grouping electric fish with catfish (fig. 7*A*) two extra steps are required (with all characters, equal weights), whereas for the clade grouping electric fish with characiforms (fig. 7*B*) three additional steps are required to break the group up (with no transitions in third positions). Although no statistical value can be attached to

Table 3
Mean Relative Synonymous Codon Usage, and Mean Frequency for Each Amino Acid

+TTT	(F)	0.85	(4.0)	TCT	(S)	0.94	(10.6)	TAT	(Y)	0.72	(4.6)	+TGT	(C)	0.94	(4.3)
+TTC	(F)	1.15		TCC	(S)	1.72		TAC	(Y)	1.28		+TGC	(C)	1.06	
+TTA	(L)	0.14	(8.7)	TCA	(S)	0.54		TAA	(*)	0.00	(0.0)	TGA	(*)	0.00	(0.0)
+TTG	(L)	0.52		TCG	(S)	0.37		TAG	(*)	0.00		TGG	(W)	1.00	(1.1)
+CTT	(L)	0.85		CCT	(P)	1.10	(4.1)	CAT	(H)	0.69	(6.1)	CGT	(R)	0.84	(2.3)
+CTC	(L)	1.53		CCC	(P)	0.75		CAC	(H)	1.31		CGC	(R)	2.44	
+CTA	(L)	0.37		CCA	(P)	0.93		CAA	(Q)	0.54	(2.3)	CGA	(R)	0.52	
+CTG	(L)	2.58		CCG	(P)	1.22		CAG	(Q)	1.46		CGG	(R)	0.70	
+ATT	(I)	0.69	(2.4)	ACT	(T)	1.27	(9.5)	AAT	(N)	0.65	(3.9)	AGT	(S)	1.19	(10.6)
+ATC	(I)	1.34		ACC	(T)	1.31		AAC	(N)	1.35		AGC	(S)	1.24	
+ATA	(I)	0.97		ACA	(T)	0.74		AAA	(K)	0.45	(4.4)	AGA	(R)	0.89	
+ATG	(M)	1.00	(2.9)	ACG	(T)	0.68		AAG	(K)	1.55		AGG	(R)	0.61	
+GTT	(V)	0.71	(8.3)	+GCT	(A)	0.98	(3.4)	GAT	(D)	0.65	(5.8)	GGT	(G)	1.02	(4.9)
+GTC	(V)	0.69		+GCC	(A)	1.83		GAC	(D)	1.35		GGC	(G)	1.51	
+GTA	(V)	0.48		+GCA	(A)	0.73		GAA	(E)	0.37	(6.4)	GGA	(G)	0.76	
+GTG	(V)	2.12		+GCG	(A)	0.46		GAG	(E)	1.63		GGG	(G)	0.71	
Column Total			26.3				27.6				33.5				23.2

NOTE.—Relative codon usage values (Sharp and Li 1987) for each triplet are followed by the mean frequency of each amino acid averaged across all sequences compared (in parentheses). Column totals are shown for amino acid frequencies only. A plus sign next to a triplet indicates a positive (>0) hydropathy index according to Kyte and Doolittle (1982).

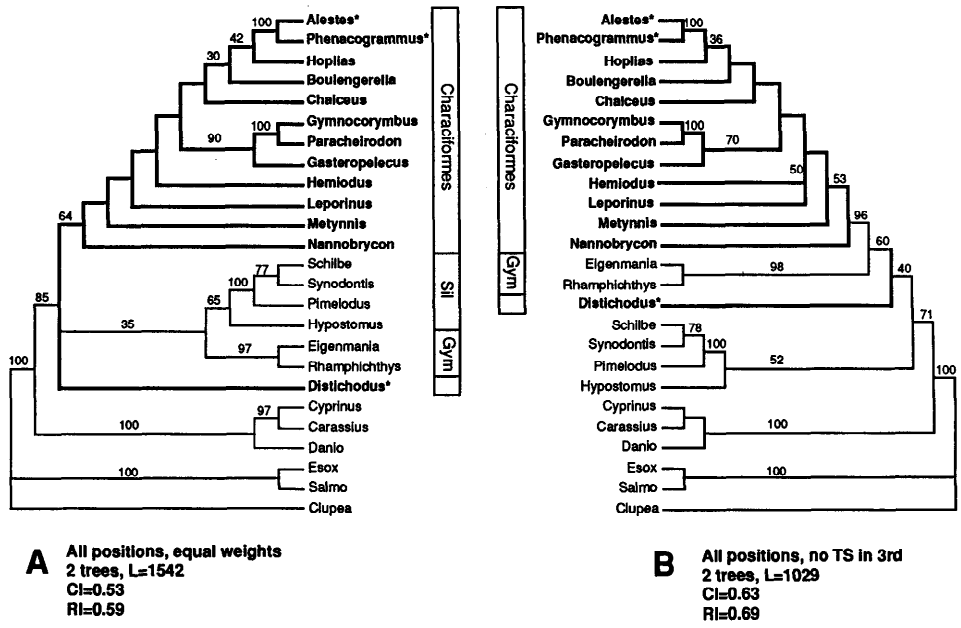

FIG. 7.—Strict consensus of phylogenetic trees obtained with parsimony. *A*, using all characters with equal weight. *B*, excluding transitions in third codon positions. Bootstrap values are shown only above branches also defined in the bootstrap majority-rule consensus tree. Bold letters and thicker branches highlight the order Characiformes, and an asterisk indicates the African characiform taxa. L is tree length, CI is the consistency index (excluding uninformative characters), and RI is the retention index (Farris 1989). Gym: order Gymnotiformes; Sil: order Siluriformes.

these decay indices, they also suggest that the grouping of electric fish with characiforms receives slightly better support than its alternative.

Neighbor-joining analyses with or without third codon positions included always grouped electric fish with characiforms. Bootstrap support (500 pseudoreplicates) was very high for Protacanthopterygii, Otophysi, Cyprinidae, Gymnotiformes, and Siluriformes (values >90) when all positions were included in the analysis. The main difference between trees including or excluding third codon positions was the placement of cyprinids and of *Distichodus*. When all positions were considered, characiform monophyly was supported with a bootstrap value of 63 and electric fish and characiforms grouped together with a bootstrap value of 42. When third positions were excluded, *Distichodus* grouped with electric fish and this clade grouped with characiforms, supported by bootstrap values of 29 and 67 respectively. Excluding third positions also had the effect of placing cyprinids as sister group of characiforms + electric fish, to the exclusion of catfish. Protein Poisson-corrected distances and Kimura distances excluding third positions resulted in the same topology. Relationships among characiform lineages were poorly supported in the NJ trees, but agreed with parsimony analyses in placing *Distochodus* at the base of characiforms, and in grouping *Alestes* + *Phenacogrammus* and *Paracheirodon* + *Gymnocorymbus* + *Gasteropelecus* with high bootstrap support.

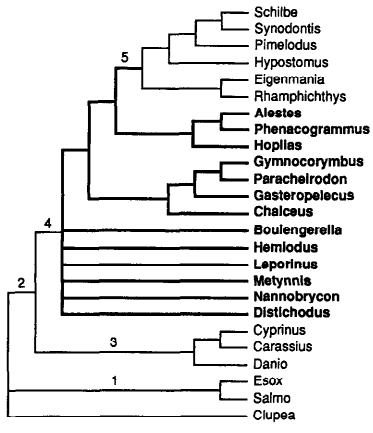

FIG. 8.—Strict consensus of three shortest trees obtained with parsimony using only third codon positions. Tree length is 853, consistency index (CI) = 0.46, and retention index (RI) = 0.54. Numbers on branches identify clades discussed in the text.

Are Third Codon Positions Phylogenetically Informative?

According to the g_1 test (Hillis and Huelsenbeck 1992), there is significant hierarchical structure at third codon positions ($g_1 = -1.102$, $P < 0.01$, for 5,000 random trees). Moreover, a heuristic search based on third positions alone recovered three most-parsimonious trees, a strict consensus of which shows much of the structure seen when the whole data set was analyzed (fig. 8), especially at the deeper branches. Also, catfish and electric fish were grouped together by third positions alone. That the most basal nodes were recovered (i.e. Protacanthopterygii, Otophysi, and Characiphysi) seemed particularly surprising.

Pairwise distances between these taxa range from 40%–60% at third codon positions (fig. 5 and table 2), well above the anticipated saturation level (Friedlander, Regier, and Mitter 1994). However, for all pairwise

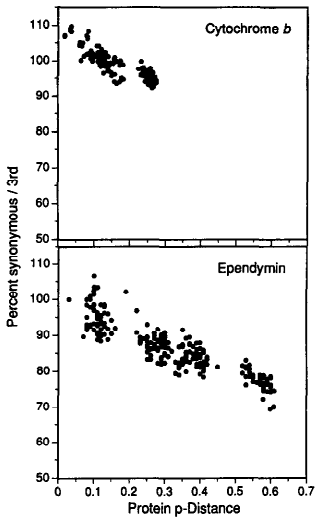

FIG. 9.—The total number of synonymous changes, expressed as a percentage of all third codon position changes (for all pairwise comparisons between taxa) is graphed as a function of p-distance (the proportion of different amino acids between two protein sequences). The top panel shows data for complete cytochrome *b* sequences of sturgeon, trout, cod, carp, loach, shiner, five cichlids, five cetaceans, zebra, and hippopotamus (from GenBank and Meyer et al., unpublished data). The bottom panel shows data for ependymin sequences and taxa included in the present study.

comparisons the proportion of nonsynonymous changes at third codon positions increased with increasing protein sequence divergence (fig. 9). Almost all changes at third codon positions are synonymous when protein divergence is low (below 20%), but synonymous changes constitute less than 80% of all changes at third codon positions when protein divergence is high (fig. 9*B*). Protein divergence between protacanthopterygians, *Clupea,* and otophysans is well over 50% (table 2), and the proportion of nonsynonymous substitutions at third codon positions among these taxa is at its maximum (fig. 9*B*). These nonsynonymous changes might retain enough phylogenetic information to resolve the deeper nodes of the tree. Indeed, 22 of 29 (75%) unambiguous changes at third codon positions reconstructed by MACCLADE on the branch leading to *Salmo* + *Esox* (fig. 8, branch 1) were nonsynonymous, as were 60% of the 17 unambiguous changes reconstructed on the branch leading to Otophysi (fig. 8, branch 2). On branches leading to Cyprinidae and Characiphysi (fig. 8, branches 3 and 4, respectively) only 6 out of 17 (35%) and 4 out of 12 (33%) changes unambiguously reconstructed were nonsynonymous, respectively. In agreement with the abovementioned base compositional bias in third codon positions of cyprinids, 10 of the 17 changes on the branch leading to Cyprinidae were G/C to T changes. In contrast, lower protein sequence divergence among the Characiphysi (characiforms, catfishes and electric fishes differ on average by 30%) may not allow accumulation of enough phylogenetically informative nonsynonymous changes at third codon positions to solve relationships at this level. All five unambiguous changes at third positions reconstructed along the branch uniting catfish and electric fish (fig. 8, branch 5) were synonymous (and three of them T-to-C transitions).

Relatively low sequence divergence at the amino acid level constrains third codon positions to accumulate mostly synonymous substitutions, which presumably do not retain phylogenetic signal. A comparison of ependymin with the highly conserved mitochondrial cytochrome *b* protein illustrates this argument further. Figure 9 shows that cytochrome *b* protein sequence divergence is small and that more than 95% of third codon position changes are silent substitutions, even when comparing such divergent taxa as fish and mammals. Cytochrome *b* has been shown to perform poorly in phylogenetic reconstruction of ancient splits (Meyer 1994b). In conclusion, third codon positions seem to retain more phylogenetic information for deep phylogenetic splits when protein sequence is poorly conserved (e.g., ependymin) than when protein divergence is low (e.g., cytochrome *b*). As a first approximation, the cut-off point for amount of amino acid sequence divergence seems to be around 30%–50%: below 30% most changes at third codon positions are synonymous and retain poor phylogenetic information, but above 50% enough nonsynonymous changes at third codon positions might provide a reliable phylogenetic signal. Friedlander, Regier, and Mitter (1994) and Soto-Adames, Robertson, and Berlocher (1994) obtained similar results concerning the information content of third codon positions for deep phylogenetic splits, but no explanation was offered in either case. At the opposite end of taxonomic divergence, phylogenetic information content of third codon positions is also significant. Among very closely related taxa, with DNA sequence divergence <5%, back-mutations have not yet accumulated to saturation levels and third codon positions retain significant phylogenetic signal (e.g., for the cytochrome *b* gene at the intraspecific level in stickleback fish, Ortí et al. 1994).

Base Composition and Phylogeny

It has been shown that tree-building methods may be unreliable when taxa differ in base composition (e.g., Saccone, Pesole, and Preparata 1989; Sidow and Wilson 1990; Lockhart et al. 1992; Hasegawa and Hashimoto 1993; Steel, Lockhart, and Penny 1993). To test whether grouping of sequences may be due to similar base compositional profiles rather than shared ancestry we applied the method suggested by Lockhart et al. (1994, p. 608). NJ trees were built using a matrix of Euclidean distances between nucleotide frequencies for each pair of taxa ($d^2 = \Sigma\ (X_{ik} - X_{jk})^2$, where X_{ik} and X_{jk} are the frequencies of nucleotides k = G, A, T, C, for taxa i and j, respectively). These "GC trees" are based only on nucleotide frequencies. Therefore, groupings found in GC trees and in trees obtained by other methods are significantly affected by base compositional biases (Lockhart et al. 1994). Trees based on Euclidean distances were computed for all codon positions together and for each codon position separately, for all sites and for parsimony informative sites only (for a total of eight trees, using the herring as an outgroup).

None of the GC trees generated recovered the monophyly of either Siluriformes, Gymnotiformes, or Characiformes, but taxa from these orders were grouped together and randomly distributed across most trees. Therefore, resolution of relationships among these orders seems not to be affected by base compositional biases. However, as expected (see codon bias above), the grouping of cyprinids (*Danio, Cyprinus,* and *Carassius*) is significantly affected by base composition. Cyprinids group together in all but two GC trees, the exceptions being when distances were based on first or second codon positions only (for all sites). The grouping of salmoniforms (*Salmo* and *Esox*) was also obtained in GC trees when only second codon positions were considered. Most remarkably, the basal nodes of the GC tree for second codon positions (for parsimony sites) are almost identical to nodes 1–4 in figure 8 (obtained by parsimony analysis on third codon positions only). Third codon position GC trees recover only branches 3 and 4 (fig. 8), but not the grouping of salmoniforms. If these basal nodes (fig. 8, branches 1–4) reflect the "true" phylogenetic history of this group of taxa (see fig. 1), what do these observations mean? Selective constraints on variation at second codon positions imposed by hydrophobicity profiles (Naylor, Collins, and Brown 1995; fig. 4) effectively limit the number of possible states, suggesting that the level of homoplasy at these sites should be high. Therefore, base compositional similarity at second codon positions may be considered a homoplasy. Likewise, phylogenetic information at third codon positions seems to be significantly affected by base compositional profiles. But the question remains whether base compositional biases (at second and, to a lesser extent, third codon positions) might reflect "true" history at this level, rather than being misleading factors for phylogenetic inference. In any case, features that are subject to selection (as, in this case, base compositional biases at second codon positions) need not have evolved more than once, and may be highly conserved (Donoghue and Sanderson 1992), thus retaining phylogenetic information.

To test whether base composition alone accounts for the grouping of Salmoniformes and the basal nodes of the tree, the LogDet transformation (Lockhart et al. 1994) was applied to the sequence data to adjust for unequal nucleotide composition. A distance matrix generated from the log determinant values was obtained using the program SPLITSTREE (Huson and Wetzel 1995) for data sets consisting of all sites or parsimony sites only, and for first and second or third codon positions only (total of four matrices). These matrices were used to calculate NJ trees, using the herring as an outgroup. These trees reflect the historic information contained in the aligned sequences without the misleading effect of base compositional biases (Lockhart et al. 1994). All NJ trees based on LogDet-transformed distances grouped *Esox* and *Salmo,* failed to recover a monophyletic Cypriniformes (*Cyprinus* falls within the catfish clade), and placed the electric fish with Characiformes. Therefore, base compositional bias is not a significant factor affecting phylogenetic analyses of this group of taxa. But these results also suggest that for particular groupings (cypriniforms, salmoniforms, and the basal nodes of the tree), base composition alone (the GC trees) might help recover rather than confound historic relationships among sequences.

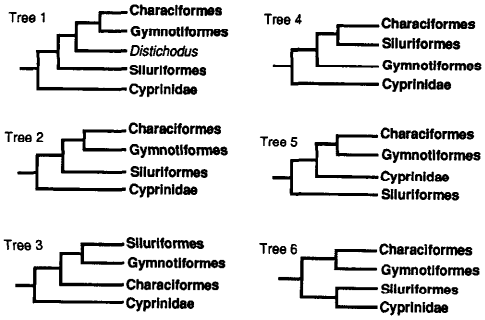

FIG. 10.—Alternative tree topologies tested by maximum likelihood and parsimony (see table 4). Branching order within each of the lineages is as shown in figure 7*B*. *Clupea, Esox,* and *Salmo* are designated outgroups. Note that tree 1 is identical to tree B in figure 7 and tree 3 to tree A (the traditional hypothesis).

Maximum Likelihood Analysis

Maximum likelihood analysis was used to evaluate the information content of third positions further and to compare alternative hypotheses (fig. 10). The rate of change at each codon position was estimated by counting the number of changes reconstructed over the shortest tree (tree B in fig. 7) using the program MACCLADE. These values were 373, 270, and 860 for first, second and third positions, respectively. They were used as auxiliary information with the input to the FASTDNAML program to activate the "categories and rates" option (Olsen et al. 1994). Five runs of the program using the jumble input option (total examined: 27,249 trees) resulted in the same best tree every time (identical to tree B in fig. 7), with a log likelihood of −6906.79. The alternative topology (fig. 7*A*) had a log likelihood of −6929.36. The same best tree (fig. 7*B*) was obtained in 3 out of 10 "jumbled" runs of FASTDNAML with only first and second positions in the data set. To evaluate the extent to which the best tree is significantly better than its alternatives, the standard errors (SE) of the differences between log likelihoods (Δl_i, Kishino and Hasegawa 1989) were computed using the program NUCML (Adachi and Hasegawa 1994; Hasegawa, Kishino, and Yano 1985). Table 4 shows tree lengths and log likelihood differences for these trees and alternative topologies (shown in fig. 10), using data sets including either all positions or only first and second positions (NUCML does not allow rate categories in the input). The differences in log likelihood between trees are not statistically significant since all upper bounds of the 95% confidence intervals (Δl_i + 1.96 [SE]) are greater than zero. According to Kishino and Hasegawa (1989), this means that none of the best trees is significantly better than the alternative hypotheses (fig. 10). However, the data set including only first and second codon positions provides somewhat better resolution among alternative trees than the one including all po-

Table 4
Lengths and Likelihoods for Trees Described in Figure 10

Tree	Parsimony: Tree Lengths (Extra Steps)			Nucleotide Maximum Likelihood				Protein Maximum Likelihood	
				All Sites		1st + 2nd			
	All Sites	NoTS in 3rd	1st + 2nd	Log Likelihood	$\Delta l_i \pm$ SE	Log Likelihood	$\Delta l_i \pm$ SE	Log Likelihood	$\Delta l_i \pm$ SE
1	1,546 (+4)	1,029	672	−7,131.8	−6.8 ± 9.0	−3,429.3	0.0	−3,611.0	0.0
2	1,547 (+5)	1,034 (+5)	676 (+4)	−7,131.1	−6.1 ± 4.7	−3,436.8	−7.5 ± 5.4	−3,614.0	−3.0 ± 3.2
3	1,542	1,037 (+8)	678 (+6)	−7,125.0	0.0	−3,439.4	−10.1 ± 7.2	−3,615.7	−4.7 ± 4.4
4	1,548 (+6)	1,034 (+5)	679 (+7)	−7,131.5	−6.6 ± 4.4	−3,441.0	−11.7 ± 6.6	−3,616.0	−5.0 ± 4.1
5	1,557 (+15)	1,040 (+11)	678 (+6)	−7,143.3	−18.3 ± 7.5	−3,441.0	−11.7 ± 6.7	−3,615.1	−4.1 ± 4.1
6	1,558 (+16)	1,046 (+17)	682 (+10)	−7,143.3	−18.3 ± 7.5	−3,441.3	−12.1 ± 6.6	−3,615.1	−4.1 ± 4.3

Note.—For parsimony searches either all characters (All Sites), transitions in third codon positions excluded (NoTS in 3rd), or first and second codon positions only (1st + 2nd) were included in the data set. Maximum likelihood analyses of nucleotide data were performed using all sites (588 sites), or first and second codon positions only (1st + 2nd: 392 sites) with the program NUCML 2.2 (with A/B = 2.30 and A/B = 1.35 respectively, model of Hasegawa, Kishino, and Yano 1985). Protein maximum likelihood analysis was performed with PROTML 2.2 ("JTT" model, 196 sites). Δl_i is the mean difference in log likelihood between the best tree and tree *i*, and SE its standard deviation.

sitions. Whereas third codon positions might retain phylogenetic signal for the most ancient branches only, first and second codon positions seem to be less "noisy" over the whole data set. For the comparison between tree 1 and tree 3, $\Delta l_{1-3} \pm$ SE is -6.8 ± 9.0 (all data) and -10.1 ± 7.2 (first and second only), the SE being larger than the difference in the first case, smaller and closer to being significant in the second case (even though it used only two thirds of all sites). Although maximum likelihood analysis also favors the grouping of electric fish with characiforms, more data are obviously necessary to determine with confidence the best phylogenetic hypothesis.

Phylogenetic Analyses of Amino Acid Sequences

A more drastic form of weighting to reduce the noise (albeit necessarily sacrificing some of the signal) of homoplasious silent sites is the use of amino acid rather than nucleotide sequences. With this approach the total number of informative sites is greatly reduced since a single amino acid replacement may be caused by one to three changes at the nucleotide level. Out of 196 amino acids in the data set 161 were variable and 130 were phylogenetically informative. The effect of this reduction is clearly seen in the results of parsimony analyses. With all 25 taxa (using the herring as the outgroup) 30 most-parsimonious trees were obtained. Their strict consensus shows the three basal groups (herring, *Salmo* + *Esox*, and Otophysi), but fails to resolve relationships among otophysan lineages. However, the consistency index of these trees is 0.75, higher than for the trees shown in figure 7, based on nucleotide sequences. *A posteriori* reweighting (Farris 1969) results in a single tree that does not support monophyly of either Characiformes or Siluriformes, but places electric fish closer to characiforms. Excluding either protacanthopterygians or the herring from the analysis improves the resolution among otophysan lineages, resulting in two or four most-parsimonious trees, respectively. Their strict consensus places electric fish as either the sister group of Characiformes, or within Characiformes, with catfish as sister group of electric fish + characiforms, and cyprinids as the basal group of otophysans. Composition of the outgroup taxa clearly affects the resulting tree topology. The bootstrap majority rule consensus tree for all 25 taxa is congruent with the tree shown in figure 7*B* (except for the relationships among characiform taxa), but bootstrap values were low.

Neighbor-joining analysis, using either Poisson-corrected p-distances (with MEGA, Kumar, Tamura, and Nei 1993) or maximum likelihood distances (JTT model in PROTML, Adachi and Hasegawa, 1994), also placed electric fish together with characiforms but identified the cyprinids rather than catfish as the sister group of Characiformes + electric fish. Bootstrap support for the branch leading to the characiform + electric fish clade was 50%. The best tree from maximum likelihood analysis (PROTML) is the tree shown in figure 7*B*, but differences between the log likelihood of this tree and alternative topologies (fig. 10, table 4) were not statistically significant, according to the test of Kishino and Hasegawa (1989). All these results coincide in suggesting that ependymin amino acid sequences contain less phylogenetic information than nucleotide sequences and are affected by the choice of outgroup, but nonetheless still support a close relationship between electric fish and characiforms.

Relationships Among Characiform Lineages

In order to further test for the effect of outgroup choice on the resolution of characiform relationships, the more distant taxa were excluded from the analysis and only catfish and electric fish were used as outgroups. Although different results were obtained for different character weighting and reconstruction methods used (fig. 11), some elements were common to all results. The

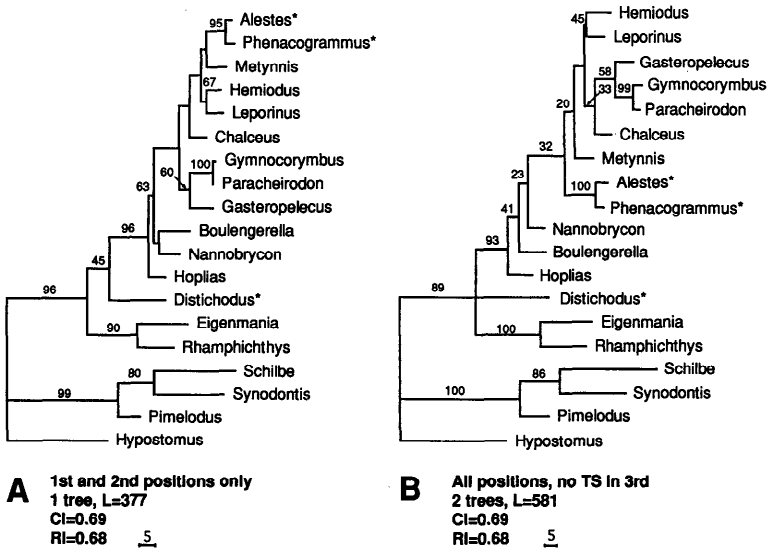

FIG. 11.—Most-parsimonious trees obtained with first and second codon positions only (*A*), or with all positions but excluding transitions in third codon positions (*B*). Branch lengths are proportional to the number of changes (scale corresponding to five changes is shown); CI is the consistency index; RI is the retention index. Bootstrap values are shown above the branches only when those branches were recovered in the majority-rule bootstrap consensus tree. African characiform taxa are indicated by an asterisk.

basal position of *Distichodus* and the grouping of *Alestes* and *Phenacogrammus* (Alestidae) and of *Paracheirodon, Gymnocorymbus,* and *Gasteropelecus* were found in all trees obtained, and supported by relatively high bootstrap values (fig. 11). These relationships were stable to outgroup choice since they were also retrieved when all 25 taxa were used (fig. 7). The position of *Chalceus* and *Metynnis* remains uncertain, but they never group together with the other taxa in the Characidae. A close relationship between *Leporinus* and *Hemiodus,* only weakly suggested in figure 7, seems to receive better support with a closer outgroup and downweighting third codon positions (fig. 11). The major discrepancy among trees involves the position of *Hoplias* and *Boulengerella.* When third codon positions (or only third position transitions) were excluded from the analysis, these taxa were no longer placed with *Alestes* + *Phenacogrammus* as a derived group within the Characiformes, but rather branched out next from *Distichodus,* at the base of the characiform clade. The same pattern is observed when amino acid sequences are used for parsimony analysis. Although no firm set of relationships can be established among characiform lineages other than those mentioned above, the monophyly of Neotropical taxa seems a very unlikely hypothesis. Under all alternative weighting strategies *Distichodus* comes out as the sister group of all other characiforms, and the Alestidae always groups among the Neotropical taxa. Forcing monophyly of Neotropical taxa results in 7, 8, and 10 extra steps when all characters were equally weighted, when transitions in third positions were excluded, and when third positions were excluded from the analysis, respectively. Mitochondrial DNA sequence evidence (Ortí and Meyer 1996) also suggests that the African and Neotropical lineages do not form reciprocally monophyletic groups.

Neighbor-joining analysis of the 19-taxon data set (with catfish as the outgroup) always resulted in a monophyletic Characiformes with *Distichodus* branching out at the base. As in parsimony analysis, by excluding third codon positions (or using protein distances) the placement of *Hoplias* and *Boulengerella* in the tree changed from being close to the Alestidae to a more basal position in the characiform clade. The grouping of *Leporinus* and *Hemiodus* was also supported, but neither the monophyly of Characidae nor that of characiforms was supported by neighbor-joining analyses. The topology of the best tree from FASTDNAML (with the categories and rates options) is the same as that shown in figure 7*B*.

Systematic and Biogeographic Implications

One of the most significant results obtained from the phylogenetic analysis of ependymin sequences is the highly supported sister group relationship of *Esox* and *Salmo* (fig. 7), corroborating, in part, the notion of Protacanthopterygii (sensu Rosen 1973, 1974) also adopted by Nelson (1994, fig. 1). Although this result was previously reported by Müller-Schmid et al. (1992), its implication for lower euteleostean systematics remained unnoticed. The superorder Protacanthopterygii, containing a diverse assemblage of basal "Division III" fishes was advanced in the seminal paper by Greenwood et al. (1966), but shortly after its inception all groups except Salmoniformes were removed (Rosen 1973). The monophyly of Salmoniformes, which included Esocoidei (pikes, mudminnows, and *Lepidogalaxias*), Argentinoidei plus Osmeroidei (smelts and their relatives), and Salmonoidei (salmonids) was proposed based on gill arch anatomy (Rosen 1974). Esocoids were later removed from the Salmoniformes and regarded as the primitive sister group of euteleosts (Fink and Weitzman 1982;

Lauder and Liem 1983; Fink 1984). Salmoniformes became coextensive with Salmonidae, and much controversy clouded the relationships among salmonids, pikes, and the other euteleosts (for a review see Fink 1984; Begle 1991, 1992; Nelson 1994). Morphological analyses have been complicated because a high proportion of characters show evolutionary losses and reductions or mosaic evolution, or exhibit a primitive condition for the euteleosts (Begle 1992; Nelson 1994). Ependymin DNA sequences have established the first molecular evidence for the monophyly of a group containing salmonids and esociforms, and hold great promise for the resolution of higher order relationships of fishes (fig. 1).

The sister group relationship of electric fish (Gymnotiformes) and Characiformes reported here constitutes a significant departure from the currently accepted hypothesis of otophysan relationships (fig. 1; Fink and Fink 1981), but had been considered the "traditional" hypothesis before 1981 (e.g., Regan 1922; Weitzman 1962; Greenwood et al. 1966; Rosen and Greenwood 1970). Gymnotiforms were then thought to be highly modified characins, albeit only based on circumstantial evidence (e.g., Mago-Leccia and Zaret 1978). The first explicit cladistic analysis of morphological characters published by Fink and Fink (1981) proposed 20 synapomorphies for the clade formed by catfish + electric fish. More recently, Dimmick and Larson (1996) presented molecular data (DNA sequences of nuclear and mitochondrially encoded rRNA genes) that support the alternative hypothesis suggested by ependymin sequences. Analyzed separately and combined, the nuclear and mitochondrial sequence data independently support the grouping of Gymnotiformes and Characiformes (Dimmick and Larson, 1996). However, in agreement with the morphological evidence, ependymin (and the nuclear and mitochondrial sequences) support the basal position of cypriniforms among otophysan lineages (fig. 7; Fink and Fink 1981).

Poor resolution among most characiform lineages (fig. 11) severely constrains interpretation of results at this level. The position of *Distichodus* as a primitive taxon among characiforms is well established by the molecular data (fig. 11) and corroborates previous morphological evidence (Fink and Fink 1981; Buckup 1991). *Distichodus* forms part of a well-defined monophyletic lineage of African characiforms composed of the families Distichodontidae and Citharinidae (Vari 1979). A close relationship of this group with the other African characiform taxa included in this study, *Alestes* and *Phenacogrammus* (family Alestidae), is not supported by either molecular (see above) or morphological evidence (Buckup 1991). Therefore, at least two levels of Afro-South American sister-group relationship are suggested, one between the distichodontids (plus citharinids) and the rest of the characiforms, and the other between alestids and a group of South American characiforms. Lundberg (1993) discussed at length the possible biogeographic implications of these phylogenetic results, and suggested a high extinction rate for African lineages as a plausible explanation.

Among the South American Characidae included in this study, a close relationship between *Paracheirodon* ("neon tetra," subfamily Cheirodontinae) and *Gymnocorymbus* ("black tetra," subfamily Tetragonopterinae) is strongly suggested (figs. 7 and 11). Tetragonopterines and cheirodontines were also suggested by Lucena (1993) to be closely related. The genera *Metynnis* ("silver dollar," subfamily Serrasalminae) and *Chalceus* (subfamily Bryconinae), usually included in the Characidae, are not shown here to form a monophyletic group with the other characids. The placement of *Metynnis* among the other putative characid taxa is equivocal (fig. 11). In an extensive survey of morphological characters, Machado-Allison (1983) presented convincing evidence for monophyly of the subfamily Serrasalminae but failed to find the sister group of this unit among characids. More recently, Lucena (1993) proposed a monophyletic group including (in addition to other taxa) serrasalmines, *Chalceus,* and *Alestes. Gasteropelecus* (family Gasteropelecidae) is shown here to have a close relationship with *Gymnocorymbus* + *Paracheirodon* to the exclusion of *Chalceus* and *Metynnis* (fig. 11). The gasteropelecids were considered to be a subfamily of the family Characidae (Weitzman 1960) but were later elevated to the rank of family by Greenwood et al. (1966). The suggestion that the family Characidae (*sensu* Greenwood et al. 1966) will undergo major taxonomic changes as phylogenetic relationships among the major lineages are established has been mentioned repeatedly (e.g., Weitzman and Fink 1983 ; Buckup 1991; Lucena 1993). Molecular systematic studies of relationships among serrasalmins, characids, and other characiform families based on mtDNA sequences may provide some fresh insights for this old question (Ortí and Meyer 1996).

Utility of Ependymin as a Phylogenetic Marker

Ependymin is a single-copy gene encoding a poorly conserved protein at the amino acid level, yet almost invariant hydrophobicity profiles, glycosylation sites, and cysteine residues (figs. 3 and 4) seem to suggest that its functional properties are retained. These features make ependymin an unlikely candidate for easy PCR amplification and direct sequencing from genomic DNA samples of distantly related taxa. Poor conservation in primary structure hinders the design of efficient primers for PCR. Indeed, the primers reported here were not suitable for PCR amplification from genomic DNA. However, as shown in the present study, RNA extraction from fresh brain tissue and reverse-transcription techniques resulted in a satisfactory approach to obtain DNA sequence data. Increased concentration of specific target cDNA obtained by reverse-transcription of brain extracts and the lower overall complexity compensates for the low binding efficiency of primers in the PCR amplification. In addition, ependymin cDNA offers the advantage of not having to sequence through intron sequences which presumably contain meager phylogenetic information for solving deep phylogenetic relationships. The major limitation imposed by this approach is the necessity to start from fresh brain tissue, precluding the use of easily collectable preserved tissues. PCR ampli-

fication from genomic DNA is only likely to work and yield substantial information with samples of closely related species (e.g., species within a genus or maybe genera within families).

Poor sequence conservation at the amino acid level is a potentially beneficial quality for a genetic marker used for reconstructing phylogenies. A wealth of substitutions among sequences should provide enough phylogenetically informative raw material for even closely related taxa. Even among distantly related taxa, poor amino acid sequence conservation should in principle attenuate the effect of saturation typically observed when the protein sequence is conserved (e.g., the mitochondrial cytochrome *b*) due to rapidly accumulating synonymous back mutations. Low structural constraints and a high proportion of sites that are free to vary should allow continuous accumulation of substitutions, even among anciently-diverged proteins. Relationships among the most divergent taxa in the present study were indeed resolved with confidence by ependymin sequences. It has been shown that third codon positions retained adequate phylogenetic information due to the high proportion of nonsynonymous substitutions that occur when protein sequence similarity is low.

Nevertheless, relationships among characiform lineages were not completely resolved by ependymin sequences in spite of seemingly appropriate amounts of variation recorded among these taxa (table 2) and relatively low levels of homoplasy in the data (consistency index = 0.7, fig. 11). Short internodes and low bootstrap values are a clear indication of poor support for most of the proposed relationships (fig. 11). Lack of resolution does not seem to be a problem derived from the pattern of molecular evolution of ependymin sequences, but might reflect a biological reality related to the radiation of characiform fishes. If diversification of the Characiformes took place over a relatively short period of time, this pattern should also be observed using other genetic markers, as seems to be the case for mtDNA sequences as well (Ortí and Meyer 1996)

In conclusion, ependymin DNA sequences should provide useful data for the study of phylogenetic relationships among basal lineages of teleost fishes. Homologous ependymin sequences have been characterized so far for Protacanthopterygii, Clupeiformes, and Otophysi. Collection of ependymin sequences from a diverse array of teleosts should be possible with the approach outlined in the present paper. At intermediate levels of divergence (i.e., among characiform families) ependymin sequences performed poorly as a phylogenetic marker. Whether the specific choice of taxa in the present study or the rapid radiation of characiform lineages may account for the poor performance observed at this level remains an open question.

Acknowledgments

This work was supported by a Doctoral Dissertation Improvement Grant to G.O. (BSR9112367) and grants to A.M. (BSR9119867, BSR9107838) and M. A. Bell (INT9117104), all from the U.S. National Science Foundation. We wish to thank A. L. Juanjo for generous laboratory assistance, J. Alves-Gomes for electric fish specimens, and M. A. Bell, D. Futuyma, R. Vari, D. Funk, S. LaForest, and I. Horovitz for discussion and two anonymous reviewers for comments on earlier versions of the manuscript. This paper was prepared in partial fulfillment of requirements for the Ph.D. in Ecology and Evolution by G.O. This is contribution 957 in Ecology and Evolution from the State University of New York at Stony Brook.

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1994. MOLPHY: a program package for molecular phylogenetics, version 2.2. The Institute of Statistical Mathematics, Tokyo.

ADAMS, D. S., and V. SHASHOUA. 1994. Cloning and sequencing the genes encoding goldfish and carp ependymin. Gene **141**:237–241.

ALVES-GOMES, J. A., G. ORTÍ, M. HAYGOOD, W. HEILIGENBERG, and A. MEYER. 1995. Phylogenetic analysis of the South American electric fishes (Order Gymnotiformes) and the evolution of their electrogenic system: a synthesis based on morphology, electrophysiology, and mitochondrial sequence data. Mol. Biol. Evol. **12**:298–318.

AVISE, J. C., J. ARNOLD, R. M. BALL, E. BERMINGHAM, T. LAMB, and J. E. NEIGEL. 1988. Intraspecific phylogeography: mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. **18**:489–522.

BEGLE, D. P. 1991. Relationships of the osmeroid fishes and the use of reductive characters in phylogenetic analysis. Syst. Zool. **41**:33–53.

BEGLE, D. P. 1992. Monophyly and relationships of the argentinoid fishes. Copeia **1992**:350–366.

BERNARDI, G., and G. BERNARDI. 1986. Compositional constraints and genome evolution. J. Mol. Evol. **24**:1–11.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution **42**:795–803.

BROWN, G. G., and M. V. SIMPSON. 1982. Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. Proc. Natl. Acad. Sci. USA **79**:3246–3250.

BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

BUCKUP, P. A. 1991. The Characidiinae: a phylogenetic study of the South American darters and their relationships with other characiform fishes. Ph.D. dissertation. The University of Michigan, Ann Arbor.

CABOT, E. L., and A. T. BECKENBACH. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. CABIOS **5**:233–234.

CARPENTER, J. 1988. Choosing among equally parsimonious cladograms. Cladistics **4**:291–296.

CATHALA, G., J.-F. SAVOURET, B. MENDEZ, B. L. WEST, M. KARIN, J. A. MARTIAL, and J. D. BAXTER. 1983. A method for isolation of intact, translationally active ribonucleic acid. DNA **2**:329–335.

DESALLE, R., T. FREEDMAN, E. M. PRAGER, and A. C. WILSON. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. **26**:157–164.

DIMMICK, W. W., and A. LARSON. 1996. A molecular and morphological perspective on the phylogenetic relationships of otophysan fishes. Mol. Phylogenet. Evol. (in press).

DONOGHUE, M. J., and M. J. SANDERSON. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. Pp. 340–368 *in* P. S. SOLTIS, D. E. SOLTIS, and J. J. DOYLE, eds., Molecular systematics of plants. Chapman and Hall, New York.

FARRIS, J. S. 1969. A successive approximations approach to character weighting. Syst. Zool. **18**:374–385.

FARRIS, J. S. 1989. The retention index and the rescaled consistency index. Cladistics **5**:417–419.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

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

FIELD, K. G., G. J. OLSEN, D. J. LANE, S. J. GIOVANNONI, M. T. GHISELIN, E. C. RAFF, N. R. PACE, and R. A. RAFF. 1988. Molecular phylogeny of the animal kingdom. Science **239**: 748–753.

FINK, S. V., and W. L. FINK. 1981. Interrelationships of the ostariophysan fishes (Teleostei). Zool. J. Linn. Soc. **72**:297–353.

FINK, W. L. 1984. Basal euteleosts: relationships. Pp. 202–206 *in* H. G. MOSER et al. , eds. Ontogeny and systematics of fishes. Am. Soc. Ichthyol. Herp. Spec. Publ. No. 1.

FINK, W. L., and S. H. WEITZMAN. 1982. Relationships of the stomiiform fishes (Teleostei), with a description of *Diplophos*. Bull. Mus. Comp. Zool. **150**:31–93.

FRIEDLANDER, T. P., J. C. REGIER, and C. MITTER. 1992. Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. Syst. Biol. **41**:483–490.

FRIEDLANDER, T. P., J. C. REGIER, and C. MITTER. 1994. Phylogenetic information content of five nuclear gene sequences in animals: initial assessment of character sets from concordance and divergence studies. Syst. Biol. **43**:511–525.

GRANTHAM, R., C. GAUTIER, M. GOUY, R. MERCIER, and A. PAVE. 1980. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. **8**:r49–r62.

GRAYBEAL, A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Syst. Biol. **43**:174–193.

GREENWOOD, P. H., D. E. ROSEN, S. H. WEITZMAN, and G. S. MYERS. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. Bull. Am. Mus. Nat. Hist. **131**:339–354.

GYLLENSTEN, U. B., and H. A. ERLICH. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQa locus. Proc. Natl. Acad. Sci. USA **85**:7652–7655.

HASEGAWA, M., and T. HASHIMOTO. 1993. Ribosomal RNA trees misleading? Nature **361**:23.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HASHIMOTO, T., Y. NAKAMURA, F. NAKAMURA, T. SHIRAKURA, J. ADACHI, N. GOTO, K.-I. OKAMOTO, and M. HASEGAWA. 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place for a mitochondria-lacking protozoan, *Giardia lamblia*. Mol. Biol. Evol. **11**:65–71.

HEDGES, S. B., K. D. MOBERG, and L. R. MAXSON. 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. Mol. Biol. Evol. **7**:607–633.

HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analysis. J. Hered. **83**:189–195.

HOFFMANN, W. 1992. Goldfish ependymins: cerebrospinal fluid proteins of meningeal origin. Prog. Brain Res. **91**:13–17.

HOFFMANN, W. 1994. Ependymins and their potential role in neuroplasticity and regeneration: calcium binding meningeal glycoproteins of the cerebrospinal fluid and extracellular matrix. Int. J. Biochem. **26**:607–619.

HUSON, D., and R. WETZEL. 1995. SPLITSTREE: An interactive macintosh application for analyzing and visualizing sequence and distance data based on the split-decomposition method, V. 1.0.41. Universitaet Bielefeld, Bielefeld, Germany.

IKEMURA, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**:13–34.

KIMURA, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA **78**:454–458.

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.

KÖNIGSTORFER, A., S. STERRER, C. ECKERSKORN, F. LOTTSPEICH, R. SCHMIDT, and W. HOFFMANN. 1989. Molecular characterization of an ependymin precursor from golfish brain. J. Neurochem. **52**:310–312.

KÖNIGSTORFER, A., S. STERRER, and W. HOFFMANN. 1989. Biosynthesis of ependymins from goldfish brain. J. Biol. Chem. **264**:13689–13692.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, Pa.

KYTE, J., and R. F. DOOLITTLE. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.

LAUDER, G. V., and K. F. LIEM. 1983. The evolution and interrelationships of the actinopterygian fishes. Bull. Mus. Comp. Zool. **150**:95–197.

LOCKHART, P. J., C. J. HOWE, D. A. BRYANT, T. J. BEANLAND, and A. W. D. LARKUM. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. J. Mol. Evol. **34**:153–162.

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. **11**:605–612.

LUCENA, C. A. S. DE 1993. Estudo filogenético da familia Characidae com uma discussao dos grupos naturais propostos (Teleostei, Ostariophysi, Characiformes). Ph.D. dissertation. Universidade de Sao Paulo, Brasil.

LUNDBERG, J. G. 1993. African-South American freshwater fish clades and continental drift: problems with a paradigm. Pp. 156–199 *in* P. GOLDBLATT, ed. Biological relationships between Africa and South America. Yale University Press, New Haven, Conn.

MACHADO-ALLISON, A. 1983. Estudios sobre la sistemática de la subfamilia Serrasalminae (Teleostei, Characidae). Parte II. Discusión sobre la condición monofilética de la subfamilia. Acta Biológica Venezuélica **11**:145–195.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: Analysis of phylogeny and character evolution, version 3.0. Sinauer Associates, Sunderland, Mass.

MAGO-LECCIA, F., and T. M. ZARET. 1978. The taxonomic status of *Rhabdolichops troscheli* (Kaup, 1856) and speculations on gymnotiform evolution. Environ. Biol. Fish. **3**: 379–384.

MEYER, A. 1993. Evolution of mitochondrial DNA of fishes. Pp. 1–38 *in* P. W. HOCHACHKA and P. MOMMSEN, eds. The

biochemistry and molecular biology of fishes. Vol 2. Elsevier, Amsterdam.

MEYER, A. 1994a. DNA technology and phylogeny of fish. Pp. 219–249 *in* A. R. BEAUMONT, ed. Genetics and evolution of aquatic organisms. Chapman and Hall, London.

MEYER, A. 1994b. Shortcomings of the cytochrome *b* gene as a molecular marker. Trends Ecol. Evol. **9**:278–280.

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

MORITZ, C., T. E. DOWLING, and W. M. BROWN. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18**: 269–292.

MÜLLER-SCHMID, A., B. GANSS, T. GORR, and W. HOFFMANN. 1993. Molecular analysis of ependymins from the cerebrospinal fluid of the orders Clupeiformes and Salmoniformes: no indication for the existence of an euteleost infradivision. J. Mol. Evol. **36**:578–585.

MÜLLER-SCHMID, A., H. RINDER, F. LOTTSPEICH, E.-M. GERTZEN, and W. HOFFMANN. 1992. Ependymins from cerebrospinal fluid of salmonid fish: gene structure and molecular characterization. Gene **118:**189–196.

MULLIS, K., F. FALOONA, S. SCHARF, S. SAIKI, G. HORN, and H. A. ERLICH. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposium on Quantitative Biology **51**:263–273.

MYERS, G. S. 1938. Freshwater fishes and West Indian zoogeography. Annu. Rep. Smithsonian Inst. **1937**: 339–364.

MYERS, G. S. 1949. Salt-tolerance of freshwater fish groups in relation to zoogeographical problems. Bijdragen tot de Dierkunde **28**:315–322.

NAYLOR, G. J. P., T. M. COLLINS, and W. M. BROWN. 1995. Hydrophobicity and phylogeny. Nature **373**:565–566.

NELSON, J. S. 1994. Fishes of the world. 3rd ed. John Wiley and Sons, N.Y.

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.

OLSEN, G. J., H. MATSUDA, R. HAGSTROM, and R. OVERBEEK. 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10**:41–48.

ORTÍ, G., M. A. BELL, T. E. REIMCHEN, and A. MEYER. 1994. Global survey of mitochondrial DNA sequences in the threespine stickleback: evidence for recent migrations. Evolution **48**:608–622.

ORTÍ, G., and A. MEYER. 1996. The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. Syst. Biol. (in press).

PARRISH, J. T. 1993. The palaeogeography of the opening South Atlantic. Pp. 8–27 *in* W. GEORGE and R. LAVOCAT, eds. The Africa-South America connection. Clarendon Press, Oxford.

PITMAN, W. C. I., S. CANDE, J. LABRECQUE, and J. PINDELL. 1993. Fragmentation of Gondwana: the separation of Africa from South America. Pp. 15–34 *in* P. GOLDBLATT, ed. Biological relationships between Africa and South America. Yale University Press, New Haven, Conn.

REGAN, C. T. 1922. The distribution of the fishes of the order Ostariophysi. Bijdragen tot de Dierkunde, Amsterdam **22**: 203–208.

RINDER, H., T. A. BAYER, E.-M. GERTZEN, and W. HOFFMANN. 1992. Molecular analysis of an ependymin gene and functional test of its promoter region by transient expression in *Brachidanio rerio*. DNA Cell Biol. **11**:425–432.

ROSEN, D. E. 1973. Interrelationships of higher euteleostean fishes. Pp. 397–513 *in* P. H. GREENWOOD, R. S. MILES, and C. PATTERSON, eds. Interrelationships of fishes. Academic Press, London.

ROSEN, D. E. 1974. Phylogeny and zoogeography of salmoniform fishes and relationships of *Lepidogalaxias salamandroides*. Bull. Am. Mus. Nat. Hist. **153**:265–326.

ROSEN, D. E. 1985. An essay on euteleostean classification. Am. Mus. Novitat. **2827**:1–57.

ROSEN, D. E., and P. H. GREENWOOD. 1970. Origin of the Weberian apparatus and the relationships of ostariophysan and gonorhynchiform fishes. Am. Mus. Novitat. **2428**:1–25.

SACCONE, C., G. PESOLE, and G. PREPARATA. 1989. DNA microenvironments and the molecular clock. J. Mol. Evol. **29**: 407–411.

SAIKI, R. K., S. SCHARF, F. FALOONA, K. B. MULLIS, G. T. HORN, H. A. ERLICH, and N. ARNHEIM. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350–1354.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–525.

SCHMIDT, J. T., and V. E. SHASHOUA. 1988. Antibodies to ependymin block the sharpening of the regenerating retinotectal projection in goldfish. Brain Res. **446**:296–284.

SHARP, P. M., and W.-H. LI. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential application. Nucleic Acids Res. **15**: 1281–1295.

SHASHOUA, V. E. 1976. Brain metabolism and the acquisition of new behaviors. I. Evidence for specific changes in the pattern of protein synthesis. Brain Res. **111**:347–364.

SHASHOUA, V. E. 1985. The role of extracellular proteins in neuroplasticity and learning. Cell. Mol. Neurobiol. **5**:183–207.

SHASHOUA, V. E. 1988. Monomeric and polymeric forms of ependymin: a brain extracellular glycoprotein implicated in memory consolidation processes. Neurochem. Res. **13**:649–655.

SHASHOUA, V. E. 1991. Ependymin, a brain extracellular glycoprotein, and CNS plasticity. Ann. NY Acad. Sci. **627**:94–114.

SIDOW, A., and K. THOMAS. 1994. A molecular evolutionary framework for eukaryotic model organisms. Curr. Biol. **4**: 596–603.

SIDOW, A., and A. C. WILSON. 1990 Compositional statistics: an improvement of evolutionary parsimony and its deep branches in the tree of life. J. Mol. Evol. **31**:51–68.

SMITH, A. B. 1989. RNA sequence data in phylogenetic reconstruction: testing the limits of its resolution. Cladistics **5**: 321–344.

SOGIN, M. L., U. EDMAN, and H. ELWOOD. 1989. A single kingdom of eukaryotes. Pp. 133–143 *in* B. FERNHOLM, K. BREMER, and H. JORNVALL, eds. The hierarchy of life. Elsevier, Amsterdam.

SOTO-ADAMES, F. N., H. M. ROBERTSON, and S. H. BERLOCHER. 1994. Phylogenetic utility of partial DNA sequences of G6*pdh* at different taxonomic levels in Hexapoda with emphasis on Diptera. Ann. Entomol. Soc. Am. **87**:723–736.

STEEL, M. A., P. J. LOCKHART, and D. PENNY. 1993. Confidence in evolutionary trees from biological sequence data. Nature **364**:440–442.

STERRER, S., A. KÖNIGSTORFER, and W. HOFFMANN. 1990. Biosynthesis and expression of ependymin homologous sequences in zebrafish brain. Neuroscience **37**:277–284.

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*. Pp. 99–117 *in* J. A. MUSICK, M. N. BRUTON, and E. K. BALON, eds. The biology of *Latimeria chalumnae* and evolution of coelacanths. Kluwer Academic Publishers, Dordrecht.

SWOFFORD, D. L. 1993. PAUP: Phylogenetic Analysis Using Parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, Ill.

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

VARI, R. P. 1979. Anatomy, relationships, and classification of the families Citharinidae and Distichodontidae (Pisces, Characoidea). Bull. Br. Mus. (Nat. Hist.), Zool. **36**:261–344.

WEITZMAN, S. H. 1960. Further notes on the relationships and classification of the South American characid fishes of the subfamily Gasteropelecinae. Stanford Ichtyol. Bull. **7**:217–239.

WEITZMAN, S. H. 1962. The osteology of *Brycon meeki*, a generalized characid fish, with an osteological definition of the family. Stanford Ichthyol. Bull. **8**:1–77.

WEITZMAN, S. H., and W. L. FINK. 1983. Relationships of the neon tetras, a group of South American freshwater fishes (Teleostei, Characidae), with comments on the phylogeny of New World Characiformes. Bull. Mus. Comp. Zool. **150**: 339–395.

WILSON, A. C., R. L. CANN, S. M. CARR et al. (11 co-authors). 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26**:375–400.

WOESE, C. R., and G. E. FOX. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **74**:5088–5090.

WOESE, C. R., O. KANDLER, and M. L. WHEELIS. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucaria. Proc. Natl. Acad. Sci. USA **87**:4576–4579.

JAN KLEIN, reviewing editor

Accepted December 19, 1995