Zootaxa 3586: 195–210 (2012)
www.mapress.com/zootaxa/

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

urn:lsid:zoobank.org:pub:66B1A0F0-5912-4C52-A9EA-D7265024064B

Phylogeny of suckers (Teleostei: Cypriniformes: Catostomidae): further evidence of relationships provided by the single-copy nuclear gene IRBP2

WEI-JEN CHEN[1] & RICHARD L. MAYDEN[2]*

[1]*Institute of Oceanography, National Taiwan University, No.1 Sec. 4 Roosevelt Rd. Taipei 10617, Taiwan.*
E-mail: wjchen.actinops@gmail.com
[2]*Department of Biology, 3507 Laclede Ave, Saint Louis University, St. Louis, Missouri, 63103 USA.*
E-mail: cypriniformes@gmail.com
[*]*Corresponding author: R.L. Mayden*

Abstract

The order Cypriniformes and family Catostomidae, the Holarctic suckers, have received considerable phylogenetic attention in recent years. These studies have provided contrasting phylogenies and classifications to historical, morphology-based phylogenetic and prephylogenetic hypotheses of relationships of species and the naturalness of hypothesized genera, tribes, and subfamilies. To date, nearly all molecular work on catostomids has been done using DNA sequence variation of mitochondrial genes. In this study, we add to our previous investigations to identify single-copy nuclear gene markers for diploid and polyploid cypriniforms, and to expand sequences of nuclear IRBP2 gene to 1,933 bp for 23 catostomid species. This effort expands our previous studies using only partial sequences of 849 bp. The extended gene fragment consists of nearly the complete gene across exon1 to exon 4 and is used in two analyses to infer phylogenetic relationships of the currently, or formerly, recognized genera, tribes, and subfamilies. One analysis includes 23 ingroup species for which the larger fragment of IRBP2 could be obtained; these taxa were also included in a second analysis of 67 samples of 52 species for the shorter fragment. As is typical of other nuclear genes examined to date for cypriniform species, variation in IRBP2 provided strong nodal support for some supra-specific groupings and species relationships. The two analyses revealed slightly different relationships, yet are largely consistent with one another. The resulting tree from variation in the shorter fragment for 52 species is somewhat inferior to the tree derived using the extended fragment in that not as many nodes were resolved, and few have strong support. Relationships from the latter analysis are, however, consistent with inferred relationships that are more robustly supported in the smaller taxon analysis using the larger fragment, lending credence to the use of more complete sequence data of genes in phylogenetic analyses. The current classification of the family (e.g., Nelson 2006) is not fully supported herein. The Ictiobinae is monophyletic, but some ambiguity exists as to relationship of this group relative to Cycleptinae and Myxocyprininae, as well as the need to recognize the latter two subfamilies. Catostominae is monophyletic. *Catostomus* is clearly not monophyletic; unnaturalness of the genus is supported herein as well as in multiple, consistently repeated and highly supported studies resolving *Deltistes*, *Chasmistes*, and *Xyrauchen* within *Catostomus.* We herein synonymize the former three genera into the latter genus; their recognition as distinct genera has been based on historical methods of classification based strictly on "distinctiveness" or anagenesis of each lineage alone and not phylogenetic relationships relative to species of *Catostomus*. The monophyly of Erimyzonini is strongly supported within the analysis of the longer sequence data set. The monophyly of Thoburniini is ambiguous, but Moxostomatini, including "*Scartomyzon*," is monophyletic in both analyses. The proposed recognition of *Scartomyzon* as a monophyletic group separate from *Moxostoma* is again falsified but with evidence from the nuclear gene IRBP2.

Key words: interphotoreceptor retinoid-binding protein gene; IRBP2, molecular systematics; polyploidy; gene duplication

Introduction

Catostomidae, Holarctic suckers, is recognized as a monophyletic group in analyses using either morphological and molecular characters, and is one of several families from the world's largest clade of freshwater fishes—the Cypriniformes. Catostomids are closely related to the clade containing three species of algae eaters (Gyrinocheilidae) plus the diverse lineage of loaches (Cobitidae and other related families). These three lineages form a major, large and diverse lineage that has been reclassified in the Superfamily Cobitoidea (Nelson 2006; Saitoh *et al.* 2006;

Accepted by L. Page: 29 Oct. 2012; published: 14 Dec 2012

Mayden & Chen 2010). Ancient suckers were once thought to be widely distributed across temperate Asia and North America (Smith 1992). The oldest fossils are known from the Eocene-Oligocene (Wilson 1977; Bruner 1991; Smith 1992). Today, extant suckers are hypothesized to include 72 species in 13 genera, and almost all species are endemic to North America south into northern Central America in Guatemala (Smith 1992; Nelson 2006). Species of this family constitute about 7% of the modern ichthyofauna in most North American freshwater ecosystems (Harris & Mayden 2001). Only one North American sucker species, *Catostomus catostomus* (Forster), has its distribution extending out of North American waters and into eastern Siberia (Nelson 2006). The highly enigmatic species *Myxocyprinus asiaticus* (Bleeker) is the only species of the family found solely outside of North America and is endemic to the Yangtze River basin of China (Smith 1992). Although suckers are not usually fished recreationally and are thus of little commercial importance, they display great diversity in endemism (Lee *et al.* 1981), morphology (Smith 1992), genomic features (Ferris 1984), and life history traits (Fuiman 1985). These fishes have attracted a great deal of attention from biologists with different research foci in biology (Harris & Mayden 2001).

In the present study, we report new primers (primarily specific to suckers) that permit the amplification and sequencing of a large gene fragment consisting of nearly the complete IRBP2 gene across exon1 to exon 4. We hope that the hypothesis(es) presented herein would be subject to testing of the 'single-copy IRBP2 gene' hypothesis in polyploid suckers but with considerably more sequence data. The second objective is to provide further evidence found in many published and ongoing studies to resolve the sister-group relationships of species of catostomids, but in this case based solely on IRBP2 gene sequence variation of the single-copy nuclear gene. Our analyses represent the most complete evaluation of relationships of species and supraspecific taxa within Catostomidae facilitated using nuclear gene variation in 23 and 52 species using long and short sequence reads, respectively. The resulting hypothesis will be compared with previous hypotheses of catostomid relationships, and more recent phylogenies based 1) solely on mitochondrial genes, specifically an ND4/ND5 gene data set (3,436 bp; Doosey *et al.* 2010) and 2) a phylogeny of moxostomatine based on cytochrome *b* and intron sequences from one of the copies of sucker growth hormone genes (Clements *et al.* 2012). Finally, the current classification of the Catostomidae (e.g., in Nelson 2006) will be evaluated and discussed.

Multiple hypotheses of relationships, homology, and nuclear genes. Given the popularity of species of catostomids across multiple disciplines, a clear taxonomic assessment leading to a natural classification of genera and species has been a long-time subject of debate. This debate is fueled due primarily to six major factors, including 1) historic studies lacking in phylogenetic argumentation; 2) a desire for some to retain such classifications for purposes of "stability;" 3) analyses combining an array of features possessing the same qualities cautioned today with regard to combining data in analyses—as in different gene trees and early methods of phylogenetic analysis (Smith 1992); 4) analysis of varied and a large number of characters using ordered character transformation series—an invalid assumption that heavily weights a resulting outcome (Smith 1992); 5) a continued recognition by some authors of three western North American genera (*Deltistes, Chasmistes, Xyrauchen)* only because of "morphological distinctiveness" and an untestable hypothesis of introgression of each species with *Catostomus*; and 6) variation in resolved relationships that may derive from the basic limitations of taxon/character sampling (Hillis & Bull 1993; Zwickl & Hillis 2002; Mayden *et al.* 2008)

Most recent molecular systematic studies of Catostomidae have been based on mitochondrial genes (Harris & Mayden 2001; Harris *et al.* 2002; Doosey *et al.* 2010), largely because of the polyploidy of these fishes and complexities of amplifying and comparing orthologous gene sequences. It has been well recognized that the exclusive use of mitochondrial genes as markers for phylogenetic reconstruction may be problematic because of inherent attributes associated with these genes or genomes relating to hybridization or introgression, independence of genes, and maternally inherited genomes (Chen *et al.* 2008; Mayden *et al.* 2009; Chen & Mayden 2010). Therefore, it is possible that a resulting "gene tree(s)" or "phylogeny" based on mitochondrial DNA sequences may not reflect a "species tree" (Mayden *et al.* 2009; Chen & Mayden 2010).

Independent evidence from different nuclear gene markers is sorely needed as independent hypotheses of relationships within the family and many other diploid or polyploid cypriniform fishes. Despite a tremendous amount of data and phylogenetic analyses, only recently have investigations accumulated from the nuclear genome for systematic and evolutionary studies in cypriniform fishes (Chen *et al.* 2008; Chen *et al.* 2009; Chen & Mayden 2009; Mayden *et al.* 2009; Mayden & Chen 2010; Palandacic *et al.* 2010; Saitoh *et al.* 2010; Tao *et al.* 2010; Tang *et al.* 2011). It has become also abundantly clear that obtaining orthologous nuclear gene sequences, especially in polyploid species, is much more difficult than sequencing and analyzing genes and has been largely avoided. Thus,

few nuclear-gene based studies on suckers were conducted (Ferris & Whitt 1978; Bart *et al.* 2010) until the development of a series of nuclear gene markers by Chen *et al.* (2008) and Chen & Mayden (2009). Even with the availability of these and other nuclear genes, because of the requisite details required in working with polyploid taxa to assure proper homologous (orthologous) gene comparisons, their application in systematic studies of Cypriniformes has been highly limited. Adding to this notable paucity of nuclear gene data is the intrinsic well documented difficulty with finding nuclear genes with the degree of variation (lineage anagenesis) needed to provided information essential as evidence for genealogical relationships among species or smaller (more recent) supra-specific groups. Nuclear genes, in general, evolve more slowly (except sections of some introns) than most mitochondrial gene regions, and as such their anagenetic changes through time have rarely been used to trace speciation during the evolution of clade.

We are aware that a species-level phylogeny would preferably be reconstructed using single-copy gene markers; however, this is difficult and represents a problematic area for inferences in the suckers because all species are tetraploid (Uyeno & Smith 1972; Ferris 1984). Many currently employed nuclear gene markers in systematic studies of fishes appear to present an additional paralogous copy (or copies) in catostomid genomes. These markers include RAG1 (recombination activation gene 1), Rhodopsin, EGR (early growth response protein) genes (Chen *et al.* 2008), and growth hormone genes (Bart *et al.* 2010). The extra copies of these gene loci can also be amplified at the same time by simply using standard primers of these markers (Chen *et al.* 2008). Thus, this necessity to obtain paralogous gene sequences of their genomes (homologous—orthologous—at the level of the analysis) is a critical aspect that requires significant attention to detail and methodological steps to conduct meaningful systematic studies using only orthologous nuclear gene sequences of polyploids for inferring species phylogeny. First, more intensive laboratory work involves cloning, a step that is obligatory if one is to resolve individual gene sequences of each paralogous copy to resolve a clear phylogenetic (gene) tree (Saitoh & Chen 2008; Saitoh *et al.* 2010) and not confuse orthologous with paralogous gene copies (not homologous at the level of the analysis). Analyses of datasets containing a mixture of orthologous and paralogous copies will yield spurious results of relationships given the differential variation, timing and dynamic processes of the gain/loss (e.g., duplication, loss, anagenesis in a lineage) of genes during the evolution of species (Chen & Mayden 2010). For instance, Bart *et al.* 2010 identified at least two distinct copies of growth hormone (GH) gene in catostomids. The homology of these copies to one another and to GH of other species of the Cypriniformes remains obscure. Nuclear genes, for the most part, are more complex and require significantly more time for sequence acquisition and alignment for research by many molecular systematists; dealing with polyploidization in taxa makes the situation much more difficult to ensure that the targeted orthologous gene fragments are amplified, sequenced, and analyzed, under the hypothesis that they are homologous at the species level.

The targeted nuclear gene marker in this study is interphotoreceptor retinoid-binding protein (IRBP) gene 2. Drawing from our earlier preliminary survey of nuclear markers (Chen *et al.* 2008), IRBP2 is a single-copy gene in the tetraploid genome of catostomid species. IRBP mediates the transfer of all-*trans* retinol and 11-*cis* retinal between the pigmented epithelium and the photoreceptors (Pepperberg *et al.* 1993). The human IRBP gene is ~ 9.5 kbp and consists of a long exon 1 plus three short exons (2 ~ 4) separated by three introns (Fong *et al.* 1990). The human IRBP exon 1 is 3051 bp but only 1194 bp for the Zebrafish (*Danio rerio*) due to a subsequent loss of a partial protein-coding region in the middle of exon 1 during the evolution of the ray-finned fishes (Rajendran *et al.* 1996; Nickerson *et al.* 2006). Because the length of this gene fragment is sufficiently large and has been previously demonstrated to be phylogenetically informative in mammals (Schneider *et al.* 1996; Stanhope *et al.* 1996; DeBry & Sagel 2001; Jansa & Weksler 2004; Gaubert & Cordeiro-Estrela 2006), IRBP exon 1 has been increasingly used as a marker for phylogenetic studies of nonmammalian vertebrates, particularly teleost fishes, including relationships within the Cypriniformes (Dettaï 2004; Chen *et al.* 2008; Dettaï & Lecointre 2008; Chen *et al.* 2009; Chen & Mayden 2009; Mayden & Chen 2010; Yang *et al.* 2012). It should be noted, however, that the teleost genome apparently contains two copies of the IRBP gene arranged head-to-tail (Nickerson *et al.* 2006), wherein the first copy of IRBP1 is without introns, and that the duplication event may have occurred early in the evolution of ray-fined fishes. The molecular marker used here and in other papers on the systematics of fishes corresponds to the previously reported IRBP gene in Zebrafish (Rajendran *et al.* 1996) or IRBP2. The latter gene is, most likely, represented ubiquitously in all teleost genomes while IRBP1 has been lost or significantly reduced in size in the genomes of some teleost lineages, such as Medaka (*Oryzias latipes*) and sticklebacks (Gasterosteidae) (Nickerson *et al.* 2006; Dettaï & Lecointre 2008).

Material and methods

Samples used. Individuals of 24 species of Catostomidae were collected for this survey, encompassing all tribes and subfamilies and representing all of the 13 currently recognized genera.

DNA data collection. Tissue extraction was performed using the Qiagen DNAeasy extraction kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Extracted DNA quantity was measured by Spectrophotometer (Eppendorf). Conditions for amplification (PCR) were as follows: GoTaq® Flexi DNA Polymerase (0.5 units) (Promega), 1x reaction buffer, 2 mM of $MgCl_2$, 200 µM of each dNTP, 0.2 µM of each primer, and 20–50 ng of genomic DNA in a 25 µl of final reaction volume. Primers used and their oligo sequences are presented in Table 1. Thermocycler conditions for PCR were: initial denaturing step at 95°C for 4 min followed by 35 cycles of 95°C (for 40 s), annealing Tm (55°C) (for 40 s), and 72°C (for 1–1.5 min. depending on size of fragments), and then a final extension step of 72°C (for 7 min) before a 4°C soak. Finally, the PCR cleanup procedure followed the AMPure magnetic bead cleanup protocol (Agencourt Bioscience Corporation) and resuspension in 30 µL of sterile water. Sequences were then determined by Macrogen Inc. (Seoul, South Korea) using ABI 3730xl analyzer (Applied Biosystems).

TABLE 1. PCR/Sequencing primer information. Reverse primers are in italics.

Primer	Location	Primer sequence (5'–3')	Source
IRBP 101F	Exon1	TCMTGGACAAYTACTGCTCACC	Chen *et al.* (2008)
IRBP 1068R	Exon1	*AGATCAKGYTGTATTCCCCACTA*	Chen *et al.* (2008)
IRBP Cat 842F	Exon1	GTTGCTAAGTCARTTAACCCCATC	This study
IRBP ex4_CatR	Exon4	*GAGMAGTGTCTGAATGGCTGATT*	This study
IRBP ex2_CatF	Exon2	CGCTTTGACATGTTTGGAGAT	This study

New primer design. Our previously published IRBP2 primers (Chen *et al*. 2008) permit the amplification and sequencing of the fragment containing only part of exon 1 (about 900 bp) for catostomids. In designing new primers to fill in the "missing" sequence portions of IRBP2, the following procedures were considered. First, a genome walking strategy (Siebert *et al.* 1995) as implemented in the Universal GenomeWalker Kit (BD Biociences) was employed to obtain the outer unknown sequence (i.e., exon1 to exon 4) of the IRBP2 gene for *Catostomus commersonii* (Lacepède) and two noncatostomid taxa: *Sewellia lineolata* (Valenciennes) and *Ischikauia steenackeri* (Sauvage). The obtained sequences, together with the complete IRBP2 sequence of *Danio rerio* (retrieved from genomic databases—Ensembl: http://www.ensembl.org/), were aligned and used as a reference template for redesigning a new set of "catostomid-specific" primers for the amplification and sequencing of a fragment consisting of the nearly complete IRBP2 gene (Table 1). An on-line tool, PRIMER 3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) was used for designing necessary primers. These new primers (IRBP Cat 842F and IRBP ex4 CatR) could work equally well for some noncatostomid taxa such as *Gyrinocheilus aymonieri* (Tirant). For someone who may be interested in obtaining a complete IRBP2 sequence for a cyprinform species in general, we suggest using the standard IRBP2 primers for cypriniforms outlined in Chen *et al*. (2008) to obtain first its exon 1 sequence, then a taxon (or group) specific forward primer (located ideally at 3' end of the exon 1) designed to work in combination with a cypriniform "generalized" IRBP ex4 CatR or user self-defined reverse primer.

Phylogenetic analysis. Parologous sequences were edited and managed using Se-Al v2.0a11 (Rambaut 1996), wherein they were initially aligned with the automatic multiple alignment program MUSCLE (Edgar 2004) using the on-line server at http://www.ebi.ac.uk/Tools/muscle/index.html, and then adjusted manually based on the inferred amino acid translation, if necessary. Regions where the amount of variation was very high and the resulting alignment would likely contain invalid assertions of homology, i.e. large insertion/deletions segments showing high dissimilarity in sequence length, were discarded for phylogenetic analyses. In total, about 17 bp from the ploy-A sequence region located in intron 3 of IRBP2 were excluded.

Descriptive statistics derived from comparing sequences were conducted using PAUP*-version 4.0b10 (Swofford 2002). For our analyses we felt it important to compile two kinds of data matrices in order to best represent the taxonomic diversity of the family and examine potential problems with taxon/character sampling

impacting final hypotheses. The first matrix consisted of sequences from IRBP2 exon 1 for each sample of species, and five outgoup species (*Gyrinocheilus aymonieri*, *Sewellia lineolata*, *Ischikauia steenackeri*, *Carassius auratus* (Linnaeus), and *Danio rerio* (Hamilton)), as well as all available catostomid sequences from GenBank for the same region (Table 2). Outgroups selected were based on recent hypotheses of relationships within cypriniform fishes (Chen *et al.* 2009; Mayden and Chen 2010) and certainly on an availability of complete sequences of IRBP2 gene (e.g. the *Danio rerio* sequence from whole genome sequences in the Ensembl) for noncatostomid cypriniforms.

TABLE 2. Taxa included in this study and accession numbers of sequences in Genbank. Classification based on Nelson (2006). Sequences obtained in this study are marked with asterisk.

Classification / Taxon	Accession no.
Outgroups	
Danio rerio (Hamilton)	Ensembl
Carassius auratus (Linnaeus)	X80802
Ischikauia steenackeri (Sauvage)	JX469994*
Gyrinocheilus aymonieri (Tirant)	JX470019*
Sewellia lineolata (Valenciennes)	JX470017*
Catostomidae	
Catostominae	
Tribe Catostomini	
Catostomus ardens Jordan & Gilbert	JX470110*
Catostomus bernardini Girard	GU939633
Catostomus cahita Siebert & Minckley	GU939634
Catostomus catostomus (Forster)	GU939635
Catostomus catostomus (Forster)	JX469996*
Catostomus clarkii Baird & Girard	GU939636
Catostomus columbianus (Eigenmann & Eigenmann)	GU939637
Catostomus columbianus (Eigenmann & Eigenmann)	JX469997*
Catostomus commersonii (Lacepède)	GU939638
Catostomus commersonii (Lacepède)	JX470018*
Catostomus discobolus Cope	GU939639
Catostomus fumeiventris Miller	GU939640
Catostomus latipinnis Baird & Girard	GU939641
Catostomus leopoldi Siebert & Minckley	GU939642
Catostomus macrocheilus Girard	GU939643
Catostomus occidentalis Ayres	GU939644
Catostomus plebeius Baird & Girard	GU939645
Catostomus plebeius Baird & Girard	JX469998*
Catostomus rimiculus Gilbert & Snyder	GU939646
Catostomus santaanae (Snyder)	GU939647
Catostomus snyderi Gilbert	GU939648
Catostomus warnerensis Snyder	GU939649
Catostomus wigginsi Herre & Brock	GU939650
Chasmistes brevirostris Cope	GU939651
Chasmistes brevirostris Cope	JX469999*
Deltistes luxatus (Cope)	GU939652
Deltistes luxatus (Cope)	JX470001*
Xyrauchen texanus (Abbott)	JX470016*

.....continued on the next page

TABLE 2. (Continued)

Classification / Taxon	Accession no.
Tribe Erimyzonini	
Erimyzon oblongus (Mitchill)	GU939653
Erimyzon oblongus (Mitchill)	JX470016*
Erimyzon sucetta (Lacepède)	GU939654
Erimyzon tenuis (Agassiz)	GU939655
Erimyzon tenuis (Agassiz)	JX470003*
Minytrema melanops (Rafinesque)	JX470006*
Tribe Moxostomatini	
Moxostoma albidum (Girard)	GU939657
Moxostoma anisurum (Rafinesque)	JX470007*
Moxostoma breviceps (Cope)	GU939659
Moxostoma breviceps (Cope)	JX470008*
Moxostoma carinatum (Cope)	GU939660
Moxostoma carinatum (Cope)	JX470009*
Moxostoma cervinum (Cope)	GU939661
Moxostoma cervinum (Cope)	JX470010*
Moxostoma collapsum (Cope)	GU939662
Moxostoma congestum (Baird & Girard)	GU939663
Moxostoma duquesnii (Lesueur)	GU939664
Moxostoma erythrurum (Rafinesque)	GU939665
Moxostoma erythrurum (Rafinesque)	JX470011*
Moxostoma hubbsi Legendre	GU939666
Moxostoma macrolepidotum (Lesueur)	GU939667
Moxostoma mascotae Regan	GU939668
Moxostoma pappillosum (Cope)	GU939669
Moxostoma pappillosum (Cope)	JX470012*
Moxostoma rupiscartes Jordan & Jenkins	GU939670
Moxostoma sp. "Sicklefin redhorse”	JX470013*
Moxostoma sp. "Sicklefin redhorse”	GU939673
Moxostoma sp. "Apalachicola redhorse"	GU939671
Moxostoma sp. "Brassy jumprock"	GU939672
Moxostoma valenciennesi Jordan	GU939674
Tribe Thoburniini	
Hypentelium nigricans (Lesueur)	JX470004*
Thoburnia atripinnis (Bailey)	GU939675
Thoburnia hamiltoni Raney & Lachner	GU939676
Thoburnia rhothoeca (Thoburn)	GU939677
Thoburnia rhothoeca (Thoburn)	JX470015*
Cycleptinae	
Cycleptus elongatus (Lesueur)	JX470000*
Ictiobinae	
Carpiodes cyprinus (Lesueur)	JX469995*
Ictiobus bubalus (Rafinesque)	JX470005*
Myxocyprininae	
Myxocyprinus asiaticus (Bleeker)	JX470014*

The second matrix included 23 of the 24 samples, and outgroup taxa identified above, that possessed sequences of the nearly complete IRBP2 gene (Table 2). In this analysis we failed to obtain sequence data from the region extending from intron 1 to exon 4 for *Catostomus ardens* Jordan & Gilbert. The sequence (from exon 1 only) of this particular taxon was not included in the second data matrix. In addition, about 165 bp between 3' end of exon 1 and 5' end of intron 1 were missing in *Moxostoma breviceps* (Cope). This taxon was included in the matrix for the analyses as missing data account for only 13% of the amplified fragment or 0.6% of the entire sequence data used in this study, and should not impact the accuracy of this phylogenetic inference. Finally, because sequence alignment from intron regions often cannot be unambiguously achieved between ingroup and outgroup taxa, intron sequences from outgroups were trimmed and treated as missing data in the data matrix.

Phylogenetic analyses were based on a partitioned Maximum Likelihood (ML) method of RAxML (Stamatakis 2006) and partitioned Bayesian approach (BA) as implemented in MrBayes 3.1.1 (Huelsenbeck & Ronquist 2001). A mixed model (with GTR+G+I nucleotide substitution model) analysis was used for the combined analyses, permitting independent estimation of individual models of nucleotide substitution for each gene partition. In this study, four partitions were assigned: 3 partitions were implemented with respect to codon positions of protein-coding region (exons); 1 partition included all intron regions.

Two independent Bayesian searches were conducted for each dataset. Four independent MCMC chains consisted of 3,000,000 replicates, sampling one tree per 100 replicates. The distribution of log-likelihood scores was examined to determine that point of stationarity for each search and to decide if additional runs were required to achieve convergence in log-likelihoods across runs or searches. Initial trees with nonstationarity in log-likelihood values were discarded, and the remaining chains of trees resulting from the convergent log likelihood scores of both independent searches were combined. These trees were used to construct a 50% majority rule consensus tree.

For the RAxML, the analyses were performed with a desktop computer using user-friendly graphical front-end software, raxmlGUI version 1.0 (Silvestro & Michalak 2011). The optimal ML tree search was conducted with 100 separate runs using the default algorithm of the program from a random starting tree for each run. The final tree was selected among suboptimal trees from each run by comparing likelihood scores under the GTR+G+I model.

Nodal support was assessed with bootstrapping (BS) (Felsenstein 1985) with the Maximum Likelihood (ML) criterion, based on 1000 pseudo-replicates or the resulting *a posteriori* probabilities from partitioned BA.

Results

Characteristics of IRBP2 gene and sequence data. Using existing and newly defined primers for IRBP2 (Table 1), we successfully amplified a complete or partially complete (at least exon1) gene fragment across a wide spectrum of catostomid diversity. In most cases, the resulting sequence profiles contained none or few sites with a mixture of double-base callings on chromatograms (indicating polymorphic sites). Observed polymorphic sites on sequences could result from either PCR/sequencing of both alleles from a heterozygous diploid individual or a PCR/sequencing of all or some of the copies of genes from a polyploid individual (Chen *et al.* 2008; Chen & Mayden 2010).

Data matrix 1, consisting of the partial length of IRBP2 exon1 sequences of 67 samples of 52 catostomid species and five outgroups, aligned over 849 bp, of which 536 were variable across taxa and 171 were parsimony-informative. When sequences were compared across ingroup taxa only, 132 sites were variable and 71 of these sites were parsimony-informative. No indels were present in the alignment from this region.

TABLE 3. IRBP2 gene structure in suckers and descriptive statistics for each gene region.

	Gene region							Total
	Exon 1	**Intron 1**	**Exon 2**	**Intron 2**	**Exon 3**	**Intron 3**	**Exon 4**	
Completeness	partial	complete	complete	complete	complete	complete	partial	partial
Length (bp)	1050	152~169	192	84~91	144	122~146	156	1906~1948
No. parsimony-informative sites	55	21	21	8	16	26	11	158
No. variable sites (in %)	125 (12%)	36 (21%)	14 (11%)	33 (36%)	30 (21%)	40 (27%)	21 (14%)	299 (15%)

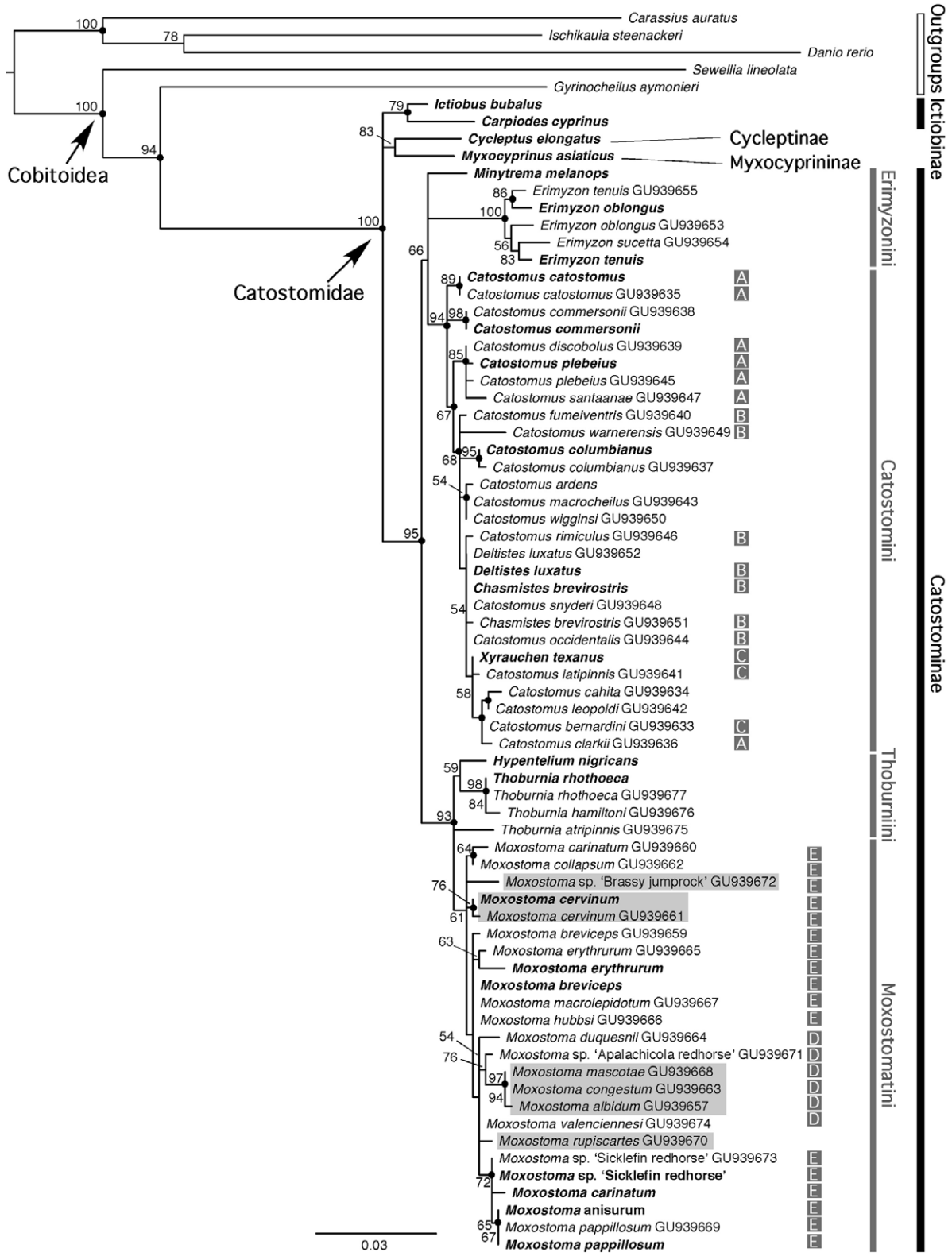

FIGURE 1. Phylogenetic tree (and/or gene tree) depicting inferred relationships of species of Catostomidae using partitioned maximum-likelihood (ML) analysis of 849 aligned nucleotides from IRBP exon 1 region in data matrix 1 (ML score—4079.257156). Branch lengths are proportional to number of substitutions under the GTR+G+I model. Numbers on branches are ML bootstrap values; those below 50% are not shown. Solid points on nodes indicate statistically robust nodes with *a posteriori* probabilities from partitioned Bayesian analysis $\geq$ 0.95. Taxa with nearly complete IRBP2 gene sequences included in this analysis (and in the analysis based on matrix 2, Fig. 2) are marked in bold. Numbers next to taxon names are Genbank accession numbers. Bars on the right indicate the classification following Nelson (2006). Solid squares A, B, C, D, and E are catostomine subclades inferred by mitochondrial genes ND4/5 analyses in Doosey *et al.* (2010). *Scartomyzon* lineages are hightlighted by gray rectangles.

Data matrix 2 consisted of the sequences of 28 taxa, including 5 outgroups, aligned over 1,933 nucleotide sites across all IRBP2 gene regions, of which 720 sites were variable and 339 sites were parsimony-informative. When sequences were compared without outgroups, 132 of 1,933 nucleotides were variable and 71 were parsimony-informative. Details of IRBP2 gene structure in catostomids and descriptive statistics for each gene region are presented in Table 3. Most of the nucleotide variability occurred at intron (especially intron 2) regions. Intron lengths were variable depending on the species. A full length IRBP2 gene sequence was obtained in this study for only *C. commersonii* and consisted of 2,226 nucleotides. The length of the reading frame was the same as the IRBP2 gene for noncatostomid cypriniforms, specifically the model species *Danio rerio* wherein the gene encodes 615 amino aides. However, IRBP2 in *Danio rerio* contains a very long (2,537 bp) intron 3 that is not found in any suckers (Table 3) or in the examined cobitoids (*Gyrinocheilus aymonieri*, *Sewellia lineolata*) where their intron 3 sequences are also available from the present study.

Inferred sister-group relationships of suckers. Inferred phylogenetic relationships (as depicted from IRBP2 gene trees) from partitioned ML and Bayesian analyses of matrices 1 and 2 resulted in hypotheses of nearly identical evolutionary relationships among species (Figs. 1 & 2, respectively). This is true except for potentially slight differences in relationships where nodal support in the tree was clearly poor (e.g., see phylogenetic position of the Erimyzonini in relation to other members of the Catostominae). However, despite the lower support for these nodes, their resolution was consistent with the alternative hypothesis. Phylogenetic evaluation of sequence variation derived from matrix 2 (Fig. 2) revealed a much more strongly supported phylogeny; 55% and 73% of nodes concerning the relationships among ingroup taxa receive high ML bootstrap values (equal to or higher than 80%) and high Bayesian posterior probabilities (equal to or higher than 0.95) (Fig. 2), respectively. This implies that longer sequence data aid in improving a global phylogenetic resolution. Indeed, the extended (or second half) part of IRBP2 sequences in this study contained about twice the amount of parsimony-informative sites as did the region from exon1 (or first half) (Table 2).

In all analyses of both reconstructions (Figs. 1 & 2), Catostomidae and subfamilies Ictiobinae and Catostominae were all resolved as monophyletic; Cycleptinae and Myxocyprininae were represented by only one species. In the larger taxon sampling analysis the relationships of the three subfamilies, Cycleptinae, Ictiobinae, and Myxocyprininae were not fully resolved but this resolution is consistent with the resolution of the monophyly of this group in Figure 2 where these subfamilies form a monophyletic group. Figure 1 simply shows a tree that is less informative as to relationships for these subfamilies. The enigmatic Asian species, *Myxocyprinus asiaticus*, which is also the only extant species from the Myxocyprininae, was resolved as the sister-group of Cycleptinae, a subfamily containing two living species surviving in the Mississippi and adjacent Gulf coastal drainages of southern United States and Mexico (Burr & Mayden 1999). This relationship was weakly supported in all analyses except in the ML analysis with matrix 1 (ML bootstrap value = 83%, Fig. 1).

Within the Catostominae, three well-supported clades were resolved, particularly in analyses of matrix 2 (Fig. 2). These three clades correspond to catostomine tribes Erimyzonini, Catostomini, and a clade grouping Thoburniini and Moxostomatini. Monophyly of the Thoburniini was never resolved. The genus *Erimyzon* was resolved as a monophyletic group (Figs. 1 & 2). Interestingly, these nuclear sequences from samples of *E. oblongus* (Mitchill) and *E. tenuis* (Agassiz), respectively, did not form monophyletic groups within each species (Fig. 1).

Regarding the genus *Catostomus*, we sampled 19 of the 24 currently recognized species. Our resulting phylogeny corroborated the clear paraphyly of *Catostomus* that results with the continued recognition of three other genera, *Chasmistes*, *Deltistes*, and *Xyrauchen*, nested within the "*Catostomus*" clade. The latter clade contains at least 4 subclades or lineages (Fig. 2). The trans-continentally distributed species, *C. catostomus*, was the sister to others of the *Catostomus* (*sensu lato* as recognized herein) clade. The widespread species, *C. commersonii* appeared in the clade following *C. catostomus* but sister to remaining species of *Catostomus* (*sensu lato* as recognized herein). *Catostomus discobolus* Cope, *C. plebeius* Baird & Girard, and *C. santaanae* (Snyder) (Figs. 1 & 2) formed a monophyletic group sister to the remaining species of *Catostomus* species and the inter-nested *Chasmistes*, *Deltistes*, and *Xyrauchen* within a monophyletic *Catostomus*. The monophyly of gene sequences within certain species and the inter-relationships among the taxa within these subclades remains unresolved. However, the placement of *Chasmistes*, *Deltistes*, and *Xyrauchen* within *Catostomus* and the closer relationship of these species to species of *Catostomus* is not controversial and robustly supported with these nuclear gene sequences in either analysis of the two matrices.

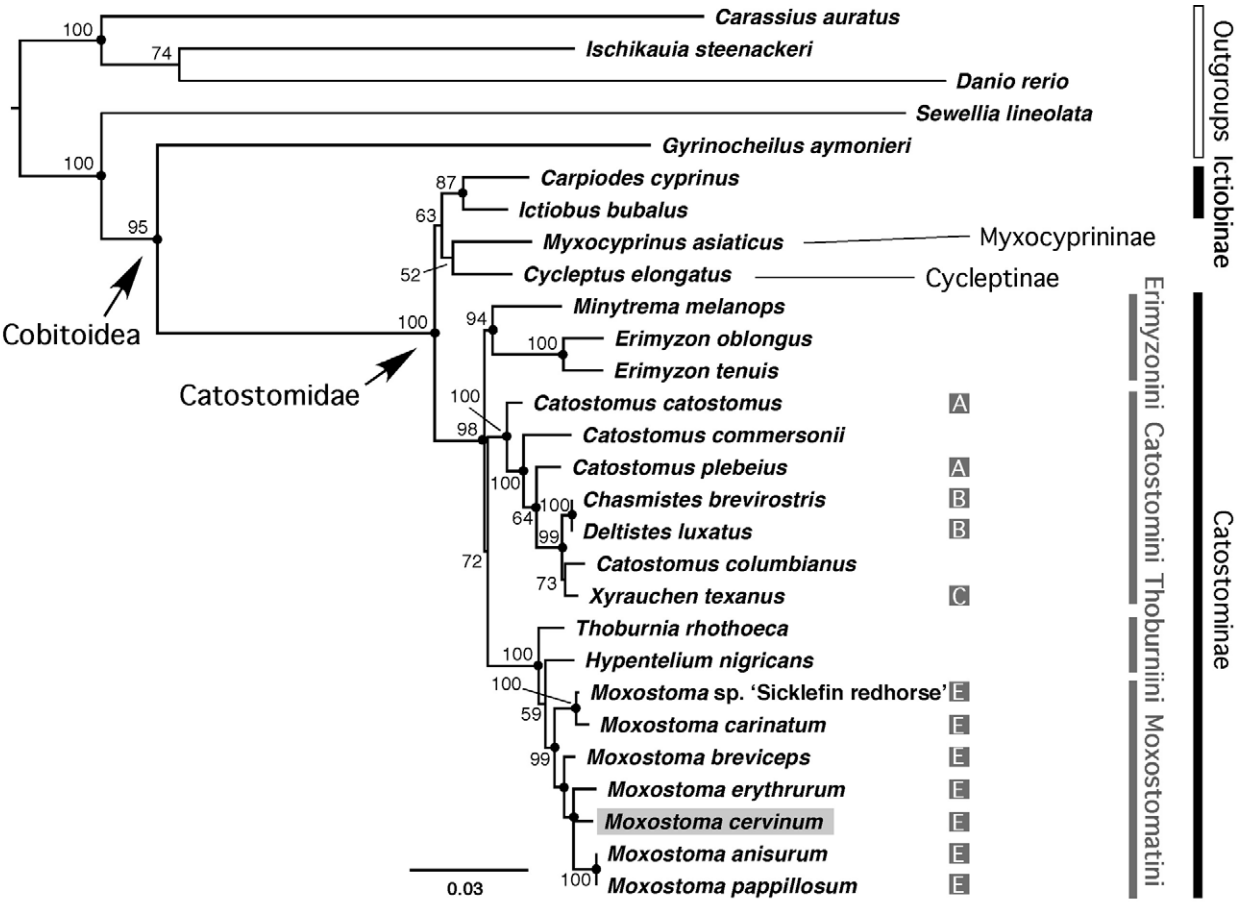

FIGURE 2. Phylogenetic tree depicting relationships of species of Catostomidae (and/or gene tree) inferred using partitioned maximum-likelihood (ML) analysis of 1,933 aligned nucleotides across all IRBP2 gene regions in data matrix 2 (ML score -8206.174368). Branch lengths are proportional to number of substitutions under the GTR+G+I model. Numbers on branches are ML bootstrap values; those below 50% are not shown. Solid points on nodes indicate statistically robust nodes with *a posteriori* probabilities from partitioned Bayesian analysis $\geq$ 0.95. Bars on right indicate the classification following Nelson (2006). Solid squares A, B, C, and E are catostomine subclades inferred by mitochondrial genes ND4/5 analyses in Doosey *et al.* (2010). *Scartomyzon* lineages are hightlighted by gray rectangles.

Moxostoma, another diverse genus within Catostomidae, includes 21 living species from which 15 described and three undescribed species were sampled. The three undescribed species are the "Sicklefin Redhorse," "Apalachicola Redhorse," and "Brassy Jumprock." The monophyly of the genus was corroborated in this study, yet interspecific relationships for *Moxostoma* were generally not well resolved, as described above for species of *Catostomus*, with the available variation in this nuclear gene, a gene that is somewhat conserved. However, the following relationships were revealed within the clade: 1) a close relationship for *M. anisurum* (Rafinesque), *M. carinatum* (Cope), *M. pappillosum* (Cope), and *M.* sp. "Sicklefin Redhorse"; and 2) a close relationship for *M. albidum* (Girard), *M. congestum* (Baird & Girard), and *M. mascotae* Regan (Figs. 1 & 2). This latter sister-species set of relationships corroborates a monophyletic Western "*Scartomyzon*" lineage; however, *Scartomyzon* (Fowler 1913) as currently recognized as a subgenus of *Moxostoma* is not monophyletic (Fig. 1). It should be noted that the two individuals of *M. carinatum* do not form a monophyletic group as the represented sequences appear in two different gene-tree clades; one sample included in the clade identified above and the other (a sequence from Genbank) as the sister-group with *M. collapsum* (Cope) (Fig. 1); however, most of the nodes supporting nonmonophyly of this gene as represented in these specimens is not well supported across all nodes within Moxostomatini (values < to 76, except for the high support for the M. *albidum*, *M. congestum*, *M. mascotae* clade).

Discussion

Single-copy IRBP2 gene in tetraploid genomes of catostomids? The evolution of nuclear genomes is inherently more complicated than that of mitochondrial genomes, with the latter having a single parental inheritance and lacking recombination (at least in animal mt-genomes). Specific characteristics such as genome duplication has diversified the genomic content that has been thought to play an important role in vertebrate evolution (Ohno 1970). In fishes, genome duplication has been hypothesized to have occurred during the early evolution of ray-finned fishes (Amores *et al.* 1998; Christoffels *et al.* 2004; Meyer & Van de Peer 2005). Ray-finned fishes (especially teleosts) usually contain more copies of many genes as compared to other vertebrates (e.g., IRBP gene focused in this study). However, as mentioned earlier, one of the critical issues in phylogenetic inference when using nuclear markers is the potential uncertainty regarding the orthology of the sequences analyzed, an issue resulting *only* in the presence of multiple copies of the genes and/or undetected paralogies (Martin & Burg 2002; Chen & Mayden 2010). Interpretations of phylogenetic results derived from nuclear gene markers should be treated with caution, like any evolutionary hypotheses, because of issues with taxon/character sampling (all data sets) and/or potential comparisons of nonhomologous or artificially identified character states (data from either molecular and morphological data sets). However, all phylogenetic hypotheses should be tested with continued investigation as to the nature of character (gene, morphology, behavior, ecology, character transformation, etc.) homology. Such continued character evaluation, expansion into new homologous and inherited characters, and the nature of reciprocal illumination (*sensu* Hennig 1966) is part of our responsibility, as part of the scientific process and as a community as a whole, to progress in knowledge acquisition related to biodiversity and evolution. The homology issue in all types of characters, specifically the nuclear genome herein, is of serious concern for all fishes having undergone gene duplication at one or more times. While mitochondrial DNA sequence data have received serious scrutiny in recent years as being problematic for phylogenetic reconstruction, equally, if not more severe concerns exist for the use of nuclear genes in inferring sister-group relations, and this is only amplified if they have had a history of gene duplication.

In this study, we used IRBP2-specific primers to avoid sampling the paralogous copy (IRBP1) of IRBP2. If by chance, IRBP1 was sequenced, it would be very easy to detect the error as the divergence between IRBP1 and 2 would be far greater than divergences observed among all teleost IRBP2 sequences because the duplication event leading to the separation of these 2 genes occurred before the diversification of all teleostean fishes. Furthermore, the presence/absence of introns is another clearly diagnostic characteristic as they can separate IRBP2 from IRBP1. The full success of amplification and sequencing of the IRBP2 fragment for all of our studied catostomid species implies that IRBP2 is, most likely, represented ubiquitously in all of their genomes without having undergone any secondary gene loss during their evolution. The occurrence of multiple, relatively recent whole-genome duplication events (polyploidy) in fishes, especially in the Cypriniformes, has been documented (Tsigenopoulos *et al.* 2002; Leggatt & Iwama 2003; Le Comber & Smith 2004). This may be of concern for the Catostomidae, as their tetraploid genomes may have originated from such an event that occurred early in the history of this lineage (Uyeno & Smith 1972; Ferris 1984; Bart *et al.* 2010). Should this hypothesis be corroborated, like many other nuclear genes, the putative duplicated copy of IRBP2 would have likely been present in the genome of their common ancestor and is retained in the genomes of extant species.

These duplicated genes (if present in the genomes) will be simultaneously amplified and sequenced when using our standard primers. However, from our resulting sequences and those retrieved from Genbank, in only a few instances (15 of 67 sequences), did polymorphic variation occur at one to several sites along a sequence (from exon 1). No sequences were detected to have an excess in the number of polymorphic sites (> 1% of sequence nucleotides according to our observations from other nuclear genes in suckers), a simple indicator of the occurrence of mistaken sequencing of all or some of the potential paralogous copies of the IRBP2 gene in these sucker species. *Ictiobus bubalus* had the highest number of polymorphic sites (9) in the IRBP exon1 sequence. The sequence divergence is far smaller than the divergence expected among all IRBP2 sequences of catostomid species. Thus, our hypothesis for the evolution of the single-copy IRBP2 gene in catostomid genomes appears to be well corroborated, making this nuclear gene effective for inferring species relationships within this family of Cypriniformes.

Systematics of suckers. (Miller 1959) depicted the first “phylogeny” of the Catostomidae. In this phylogeny, the Cycleptinae (*Cycleptus* plus *Myxocyprinus*) was the “sister-group” to all other Catostomidae. The second

subfamily, Ictiobinae, was placed at an intermediate position in the tree, sister to the last subfamily, Catostominae. Moreover, three earlier tribal designations for the Catostominae were confirmed with Catostomini and Moxostomatini being more closely related to one another than to the Erimyzonini. It should be noted that in his discussion on relationships, Miller (1959) thought that the Cycleptinae should be divided into two subfamilies but only to be consistent with the disjunct distributions of *Cycleptus* (North America) and *Myxocyprinus* (China). This opinion on the classification was adopted in later studies (see below).

Smith (1992) provided the first comprehensive analysis of catostomid relationships based on 64 taxa and 157 morphological, biochemical, and early life history transformation series. This analysis, however, was conducted using *a priori* ordering of some characters, an assumption that heavily influences the final outcome as these are considered "known evolutionary pathways for character transformations." This phylogeny was different from that presented by Miller (1959) in two ways: the Ictiobinae was resolved as the basal-most lineage, and Cycleptinae (*Cycleptus* plus *Myxocyprinus*) was the sister taxon of the Catostominae, the latter consisting of only two tribal designations: Catostomini and Moxostomatini (including the previously defined Erimyzonini).

Recent molecular systematic studies of the Catostomidae have relied largely on mitochondrial gene sequence variation. Harris and Mayden (2001) examined phylogenetic relationships of major clades of catostomids inferred from ribosomal DNA sequences. Harris *et al.* (2002) further investigated phylogenetic relationships of *Moxostoma* and *Scartomyzon* within the subfamily Catostominae based on mitochondrial cytochrome *b* sequence data. These authors confirmed the monophyly of previously defined groups except for the Cycleptinae and Erimyzonini. Their phylogenetic inferences resolved relationships necessitating several changes to the classification. This included the formation of the new subfamily Myxocyprininae, containing *Myxocyprinus* from China; restriction of the Cycleptinae to the two species of *Cycleptus* from North America; restriction of the tribe Moxostomatini to *Moxostoma* and *Scartomyzon* (currently valid as *Moxostoma*); *Erimyzon* and *Minytrema* as *incertae sedis* within the Catostominae; and resurrection of the tribe Thoburniini, containing *Thoburnia* and expanded to include *Hypentelium*. Nelson (2006) followed the classification of Harris and Mayden (2001) but retained the tribal designation of the Erimyzonini. In addition, as *Scartomyzon* was never resolved as monophyletic, but was always recovered as a polyphyletic group embedded within *Moxostoma*, *Scartomyzon* was suggested to be synonymized into the genus *Moxostoma* (Nelson 2006; Clements *et al.* 2012).

Saitoh *et al.* (2006) inferred the interrelationships of the major groups of the entire Cypriniformes using whole mitochondrial genomic data and resolved a new phylogenetic hypothesis for the relationships among four catostomid subfamilies as follows: ((Myxocyprininae, (Cycleptinae, Ictiobinae)), Catostominae). Within the Catostominae, two reciprocally monophyletic groups, *Catostomus* plus *Minytrema* and *Moxostoma* plus *Hypentelium* were well supported. Finally, hypotheses on the evolutionary relationships of catostomids was further examined by Doosey *et al.* (2010) using ND4/5 gene sequences (3,436 nucleotides) of all 13 genera and 60 species. Their analysis provided evidence for: monophyly of four subfamilies; another new hypothesis (but not well supported) on interfamilial relationships as (((Myxocyprininae, Ictiobinae), Cycleptinae), Catostominae); strong support for recognizing four catostominine tribal designations including the Erimyzonini; several well supported clades nested within Catostomini (clades A, B, and C) and Moxostomatini (clades D and E).

Analyses using IRBP2 sequence variation in this study corroborate the monophyly of two of the four catostomid subfamilies (Ictiobinae, Catostominae) (monophyly of Myxocyprininae and Cycleptinae was not tested as only one species of the former subfamily exists for sequencing and we only sampled one species of the latter subfamily). Furthermore, analysis of IRBP2 sequence variation also corroborated the monophyly of, three of the four tribal designations. The tribe Thoburniini was resolved as an unnatural group. The hypothesis of a close evolutionary affinity among Myxocyprininae, Ictiobinae, and Cycleptinae as proposed using either mt-genome or long mt sequence data from ND4/5 is corroborated by these nuclear data, but only in the reduced taxon data set (relationships depicted in Fig. 1 are consistent with this but are not resolved). Interestingly, the sister-group relationship of *Cycleptus* and *Myxocyprinus* (Miller 1959; Smith 1992) is rediscovered, supporting a natural grouping for the traditionally defined subfamily Cycleptinae resurrected by Harris and Mayden (2001).

With respect to the interrelationships of the four tribes within the Catostominae, support for a clade inclusive of all members of Thoburniini and Moxostomatini is strongly supported and this evidence has been recurrently found in all previous studies (including analyses of morphological and nuclear gene (growth hormone intron) data (Clements *et al.* 2012). *Thoburnia* and *Hypentelium* remain *incertae sedis* within the clade, and may represent the sister group to the Catostomini. The Erimyzonini, while largely inconclusive as to sister-group relationships in the

larger data set with limited sequences (potential character sampling problem), is strongly supported as a clade when using matrix 2, containing more complete sequences of IRPB2 (potentially rectifying the character sampling problem). This tribe is the sister clade to other Catostomine (Fig. 2). This particular sister-group relationship was also revealed using mitochondrial ND4/5 analyses (Doosey *et al.* 2010). However, multiple subclades (notably, subclades A, B and C; Figs. 1 & 2) resolved using mitochondrial ND4/5 data (Doosey *et al.* 2010) for the Catostomini and Moxostomatini are not resolved as monophyletic herein (Figs. 1 & 2). The subclade D from Moxostomatini is recovered as monophyletic from our analysis (node support is weak; 54% from MLBS) (Fig. 1). This clade includes a strongly supported Western "*Scartomyzon"* clade (*M. mascotae, M. congestum, M. albidum* sampled in this study). *Moxostoma* sp. 'Apalachicola redhorse' is not part of the Eastern "*Scartomyzon"* lineages but represents the sister-group of the latter clade with moderate support (Fig. 1). The monophyly for the Western "*Scartomyzon"* is also found from growth hormone intron gene sequences and the resulting tree (Clements *et al.* 2012). This clade, however, is not recovered with cytochrome *b* variation in Clements *et al.* (2012)'s study or in any other published analyses derived from mitochondrial gene sequence variation.

Importantly, informative biological classifications are those that are consistent with resolved phylogenetic relationships. Herein, the currently recognized taxonomy and classification of species at intratribal level for two different resulting groups is inconsistent with our phylogenetic inferences. In both cases no existing study provides phylogenetic evidence (synapomorphies) to contradict our inferences, demanding changes in classification. First, the monophyly of *Catostomus* is falsified herein as our analyses clearly identify the genus *Catostomus* as paraphyletic with respect to *Deltistes*, *Chasmistes*, and *Xyrauchen.* The proposed recognition of *Deltistes*, *Chasmistes*, and *Xyrauchen* is falsified. Second, *Moxostoma*, when inclusive of *Scartomyzon*, is polyphyletic. The proposed recognition of *Scartomyzon* as monophyletic group, independent of *Moxostoma* is falsified herein.

Smith (1992) resolved two reciprocally monophyletic groups within the Catostomini: one containing species of *Catostomus*, and the other involving species from *Xyrauchen, Deltistes*, and *Chasmistes*, with the latter two genera more closely related to one another. It has also been hypothesized that recurrent hybridization among certain *Catostomus, Deltistes*, and *Chasmistes* may have occurred in the history of these lineages (Miller & Smith 1981; Smith 1992; Harris *et al.* 2002). However, this is not a testable hypothesis. The most parsimonious explanation for the genetic similarities of the four putative genera (*Catostomus, Xyrauchen, Deltistes, Chasmistes*) is descent from a common ancestor, as demonstrated in our analyses. Hypotheses of postdivergence hybridization or intergradation used to explain genetic similarities among these putative genera are ad hoc stories that cannot be tested. To provide evidence for the spread of genomes across species after their divergence, the genetic constitution of each species at the time of speciation would have to be known. Also, this hypothesis assumes that mixed alleles between taxa cannot be explained by the retention of genetic variability in the lineage. Shared alleles between taxa, even if not sister taxa, does not corroborate a hypothesis of gene exchange until their presence in a taxon can be demonstrated to be the result of an active process and not historical legacy. The most parsimonious explanation of derived genetic similarity is based on synapomorphic characters, and any other more complex explanation for this similarity, like the long-thought idea of widespread genetic exchange across taxa, represents a series of declarations that are less parsimonious or cannot be tested.

Theoretically, hybridization may make evolutionary relationships more complicated to be inferred. However, congruent results of the nonmonophyletic nature of *Catostomus* in recent molecular analyses using independent gene markers and the observation of only very small amounts of nucleotide divergence in the sequences among *Xyrauchen, Deltistes, Chasmistes,* and a few species of *Catostomus* suggest that *Xyrauchen* Eigenmann & Kirsch, *Deltistes* Seale, and *Chasmistes* Jordan should be synonymized with *Catostomus* Lesueur. Certainly, usage of more powerful markers such as microsatellites may help to better understand what has been proposed to be a complex evolutionary history for species of *Catostomus*. However, the previously reported complexities in the evolution of these species were originally set forth without any phylogenetic evaluation as to the origins of the genomes and such complexity is not a testable hypothesis today. Microsatellite variation may aid in corroborating/falsifying hypotheses of gene flow amongst species, but only when evaluated in a phylogenetic context. The existence of pleisiomorphic alleles in any species following the evolution of such alleles in the evolutionary relationships of all species above a node does not constitute gene flow, but is rather equally interpreted as retention of a plesiomorphic allele and incomplete lineage (gene) sorting.

The classification of these three genera independent of *Catostomus* is the result of historical methods of taxonomy and classification wherein the taxonomic rank was highly correlated with degree (as interpreted by a

researcher) of morphological divergence and not sister-group relationships. Morphological divergences can range from great to small across taxa, and in earlier days was thought to be tightly correlated with the age of divergence of a group. Thus, high levels of divergence, in the absence of phylogenetic relationships, would indicate greater age than hypothesized lower taxonomic ranks. This method of classification is not being criticized herein as the philosophy and methods of phylogenetic systematics postdated many of these taxonomic decisions. However, strong independent evidence must be provided to support the hypothesis of these species remaining in different genera from *Catostomus* other than simply arguing that the relationships are confused by some hypothesized hybridization between unknown taxa at an unknown time in the evolution of *Catostomus*. Hybridization is a relatively common phenomenon among cypriniforms and in the vast majority of other instances it has not continued to confuse researchers in resolving phylogenetic relationships using multiple character sets. No data support these genera as forming monophyletic groups separate from *Catostomus* other than the argument that they have been involved in hybridization events; even if the hypothesis of hybridization continues to exist, there are no data to support such a claim; all molecular data support their close relationship within *Catostomus*. Perhaps further, more refined analyses of morphological data are necessary to corroborate a hypothesis of hybridization and their "true" sister-group relationships. However, evidence for genetic control of specific morphological characters is virtually absent, making such evaluations at this time moot.

Finally, two species of *Erimyzon* represented by more than one specimen were resolved as "nonmonophyletic" (Fig. 1). One potential explanation is that the sequence downloaded from Genbank is from a misidentified specimen. Another potential explanation for these unexpected results is that the specimen that is identified as one species based on morphological characters is actually carrying nuclear alleles of another species either through some type of incomplete lineage sorting or lateral gene exchange. Studies of variability in species of Erimyzon clearly needed to resolve questions of this nature. The latter explanation, however, does not seem likely because the "mistaken" sequences should be identical or nearly identical to sequences from one of three currently recognized species. Yet, our analyses of five specimens of *Erimyzon* from three species resulted in five distinct lineages (average *p* distance = 0.0093 based on IRBP2 exon1 sequences). The sequence divergence among the samples is greater than that observed in other species divergence comparisons in catostomids (average *p* distance = 0.0069 and 0.0058 for *Catostomus* spp. and *Moxostoma* spp., respectively) (we do not argue in any way that species as lineages must have a certain degree of divergence to be "real"). Given that the genus and species have not been thoroughly examined for more detailed morphological variation (including live colors) and molecular variation, this variation may support a hypothesis to be tested of the presence of undiscovered lineages within *Erimyzon.* Examination of the voucher specimen for the Genbank sequence is warranted and a careful systematic study of all of these species is needed before this hypothesis can be corroboration or falsified.

Acknowledgements

This research was supported, in part, by grants from the USA National Science Foundation (No. EF-0431326, DEB-0817027, DBI-0956370, DEB-1021840, DEB 1022720) and the William S. Barnickle endowment (to RLM) and was completed as part of an international Cypriniformes Tree of Life initiative (www.cypriniformes.org). W.-J. Chen received grant support (NSC 99-2611-M-002-001-MY2) from the National Science Council of Taiwan to conduct part of this study.

References

Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., Westerfield, M., Ekker, M., & Postlethwait, J.H. (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science,* 282, 1711–1714.

Bart, H.L., Reneau, P.C., Doosey, M.H. & Bell, C.D. (2010) Evolutionary divergence of duplicate copies of the growth hormone gene in suckers (Actinopterygii: Catostomidae). *International Journal of Molecular Sciences,* 11, 1090–1102.

Bruner, J.C. (1991) Comments on the genus *Amyzon* (family Catostomidae). *Journal of Paleontology,* 65, 678–686.

Burr, B.M. & Mayden, R.L. (1999) A new species of *Cycleptus* (Cypriniformes: Catostomidae) from Gulf Slope drainages of Alabama, Mississippi, and Louisiana, with a review of distribution, biology and conservation status of the genus. *Bulletin Alabama Museum of Natural History,* 20, 19–57.

Chen, W.-J., Lheknim, V. & Mayden, R.L. (2009) Molecular phylogeny of the Cobitoidea (Teleostei: Cypriniformes) revisited:

position of enigmatic loach *Ellopostoma* Vaillant resolved with six nuclear genes. *Journal of Fish Biology,* 75, 2197–2208.
Chen, W.-J. & Mayden, R.L. (2009) Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the World's largest clade of freshwater fishes: further evidence from six nuclear genes. *Molecular Phylogenetics and Evolution,* 52, 544–549.
Chen, W.-J. & Mayden, R.L. (2010) A phylogenomic perspective on the new era of ichthyology. *BioScience,* 60, 421–432.
Chen, W.-J., Miya, M., Saitoh, K. & Mayden, R.L. (2008) Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: the order Cypriniformes (Ostariophysi) as a case study. *Gene,* 423, 125–134.
Christoffels, A., Koh, E.G., Chia, J.M., Brenner, S., Aparicio, S. & Venkatesh, B. (2004) *Fugu* genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Molecular Biology and Evolution,* 21, 1146–1151.
Clements, M.D., Bart, H.L. & Hurley, D.L. (2012) A different perspective on the phylogenetic relationships of the Moxostomatini (Cypriniformes: Catostomidae) based on cytochrome-b and Growth Hormone intron sequences. *Molecular Phylogenetics and Evolution,* 63, 159–167.
DeBry, R.W. & Sagel, R.M. (2001) Phylogeny of Rodentia (Mammalia) inferred from the nuclear-encoded gene IRBP. *Molecular Phylogenetics and Evolution,* 19, 290–301.
Dettaï, A. (2004) *La phylogénie des Acanthomorpha (Teleostei) inférée par l'étude de la congruence taxinomique,PhD Dissertation.* Université Paris 6, Paris, 347 pp.
Dettaï, A. & Lecointre, G. (2008) New insights into the organization and evolution of vertebrate IRBP genes and utility of IRBP gene sequences for the phylogenetic study of the Acanthomorpha (Actinopterygii : Teleostei). *Molecular Phylogenetics and Evolution,* 48, 258–269.
Doosey, M.H., Bart, H.L., Saitoh, K. & Miya, M. (2010) Phylogenetic relationships of catostomid fishes (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. *Molecular Phylogenetics and Evolution,* 54, 1028–1034.
Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32, 1792–1797.
Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution,* 39, 783–791.
Ferris, S.D. (1984) Tetraploidy and the evolution of the catostomid fishes. *In:* Turner, B. J. (Ed), *Evolutionary Genetics of Fishes*. Plenum, New York, pp. 55–93.
Ferris, S.D. & Whitt, G. S. (1978) Phylogeny of tetraploid catostomid fishes based on loss of duplicate gene-expression. *Systematic Zoology,* 27, 189–206.
Fong, S.L., Fong, W. B., Morris, T. A., Kedzie, K. M. & Bridges, C. D. (1990) Characterization and comparative structural features of the gene for human interstitial retinol-binding protein. *Journal of Biological Chemistry,* 265, 3648–3653.
Fowler, H.W. (1913) Notes on catostomid fishes. *Proceedings of the Academy of Natural Sciences Philadelphia,* 6558, 45–60.
Fuiman, L.A. (1985) Contributions of developmental characters to a phylogeny of catostomid fishes, with comments on heterochrony. *Copeia,* 1985, 833–846.
Gaubert, P. & Cordeiro-Estrela, P. (2006) Phylogenetic systematics and tempo of evolution of the Viverrinae (Mammalia, Carnivora, Viverridae) within feliformians: Implications for faunal exchanges between Asia and Africa. *Molecular Phylogenetics and Evolution,* 41, 266–278.
Harris, P.M. & Mayden, R.L. (2001) Phylogenetic relationships of major clades of Catostomidae (Teleostei : Cypriniformes) as inferred from mitochondrial SSU and LSU rDNA sequences. *Molecular Phylogenetics and Evolution,* 20, 225–237.
Harris, P.M., Mayden, R.L., Perez, H.S.E. & De Leon, F.G. (2002) Phylogenetic relationships of *Moxostoma* and *Scartomyzon* (Catostomidae) based on mitochondrial cytochrome b sequence data. *Journal of Fish Biology,* 61, 1433–1452.
Hennig, W. (1966) *Phylogenetic Systematics*. University of Illinois Press, Urbana, Illinois, 263 pp.
Hillis, D.M. & Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology,* 42, 182–192.
Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES. Bayesian inference of phylogeny. *Bioinformatics,* 17, 754–755.
Jansa, S.A. & Weksler, M. (2004) Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. *Molecular Phylogenetics and Evolution,* 31, 256–276.
Schneider, H., Sampaio, I., Harada, M.L., Barroso, C.M., Schneider, M.P., Czelusniak, J. & Goodman, M. (1996) Molecular phylogeny of the New World monkeys (Platyrrhini, Primates) based on two unlinked nuclear genes: IRBP intron 1 and ε-globin sequences. *American Journal of Physical Anthropology,* 100, 153–179.
Le Comber, S.C. & Smith, C. (2004) Polyploidy in fishes: patterns and processes. *Biological Journal of the Linnean Society,* 82, 431–442.
Lee, D.S., Gilbert, C.R., Hocutt, C.H., Jenkins, R.E., McAllister, D.E. & Stauffer, J.R., Jr. (1981) *Atlas of North American freshwater fishes*. North Carolina Museum of Natural History, Raleigh, NC, 867 pp.
Leggatt, R.A. & Iwama, G.K. (2003) Occurrence of polyploidy in fishes. *Reviews in Fish Biology and Fisheries,* 13, 237–246.
Martin, A.P. & Burg, T.M. (2002) Perils of paralogy: using HSP70 genes for Inferring organismal phylogenies. *Systematic Biology,* 51, 570–587.
Mayden, R.L. & Chen, W.-J. (2010) The world's smallest vertebrate species of the genus *Paedocypris*: A new family of freshwater fishes and the sister group to the world's most diverse clade of freshwater fishes (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution,* 57, 152–175.
Mayden, R.L., Chen, W.-J., Bart, H. L., Doosey, M. H., Simons, A. M., Tang, K. L., Wood, R. M., Agnew, M. K., Yang, L., Hirt, M. V., Clements, M. D., Saitoh, K., Sado, T., Miya, M. & Nishida, M. (2009) Reconstructing the phylogenetic relationships of the Earth's most diverse clade of freshwater fishes—Order Cypriniformes (Actinopterygii: Ostariophysi): a case study using multiple nuclear loci and the mitochondrial genome. *Molecular Phylogenetics and Evolution,* 51, 500–514.
Mayden, R.L., Tang, K.L., Wood, R.M., Chen, W.-J., Agnew, M.K., Conway, K.W., Yang, L., Li, J., Wang, X., Saitoh, K., Miya, M., He, S., Liu, H., Chen,Y. & Nishida, M. (2008) Inferring the Tree of Life of the order Cypriniformes, the earth's most diverse

clade of freshwater fishes: implications of varied taxon and character sampling. *Journal of Systematics and Evolution*, 46, 424–438.
Meyer, A. & Van de Peer, Y. (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays*, 27, 937–945.
Miller, R.R. (1959) Origin and affinities of the freshwater fish fauna of western North America. *In:* Hubbs, C.L. (Ed), *Zoogeography.* American Association for the Advancement of Science, Washington D.C., pp. 187–222.
Miller, R.R. & Smith, G.R. (1981) Distribution and evolution of *Chasmistes* (Pises: Catostomidae) in western North America. *Occasional Papers of the Museum of Zoology University of Michigan*, 696, 1–46.
Nelson, J.S. (2006) *Fishes of the World, 4th Edition.* John Wiley & Sons, Inc., Hoboken, NJ, 601 pp.
Nickerson, J.M., Frey, R.A., Ciavatta, V.T. & Stenkamp, D.L. (2006) Interphotoreceptor retinoid-binding protein gene structure in tetrapods and teleost fish. *Molecular Vision*, 12, 1565–1585.
Ohno, S. (1970) *Evolution by gene duplication*. Springer-Verlag, New York, 160 pp.
Palandacic, A., Zupancic, P. & Snoj, A. (2010) Revised classification of former genus *Phoxinellus* using nuclear DNA sequences. *Biochemical Systematics and Ecology*, 38, 1069–1073.
Pepperberg, D.R., Okajima, T.L., Wiggert, B., Ripps, H., Crouch, R.K. & Chader, G.J. (1993) Interphotoreceptor retinoid-binding protein (IRBP). Molecular biology and physiological role in the visual cycle of rhodopsin. *Molecular Neurobiology*, 7, 61–85.
Rajendran, R.R., Van Niel, E.E., Stenkamp, D.L., Cunningham, L.L., Raymond, P.A. & Gonzalez-Fernandez, F. (1996) Zebrafish interphotoreceptor retinoid-binding protein: differential circadian expression among cone subtypes. *The Journal of Experimental Biology*, 199, 2775–2787.
Rambaut, A. (1996) *Sequence alignment editor version 1.0 α1*, University of Oxford, Oxford, UK. Available from http://evolve.zoo.ox.ac.uk/Se-Al/Se-Al.html.
Saitoh, K. & Chen, W.-J. (2008) Reducing cloning artifacts for recovery of allelic sequences by T7 endonuclease I cleavage and single re-extension of PCR products — a benchmark. *Gene*, 423, 92–95.
Saitoh, K., Chen, W.-J. & Mayden, R.L. (2010) Extensive hybridization and tetrapolyploidy in spined loach fish. *Molecular Phylogenetics and Evolution*, 56, 1001–1010.
Saitoh, K., Sado, T., Mayden, R.L., Hanzawa, N., Nakamura, K., Nishida, M. & Miya, M. (2006) Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): the first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. *Journal of Molecular Evolution*, 63, 826–841.
Schneider, C.J., Cunningham, M., Moritz, C. (1998) Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics rainforests of Australia. *Molecular Ecology*, 7, 487–498.
Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. & Lukyanov, S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research*, 23, 1087–1088.
Silvestro, D. & Michalak, I. (2011) RaxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution*, http://dx.doi.org/10.1007/s13127-13011-10056-13120.
Smith, G.R. (1992) Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America and Asia. *In:* Mayden, R.L. (Ed), *Systematics, historical ecology and North American freshwater fishes.* Standford University Press, Stanford, pp. 778–813.
Stamatakis, A. (2006) RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics*, 22, 2688–2690.
Stanhope, M.J., Smith, M.R., Waddell, V.G., Porter, C.A., Shivji, M.S. & Goodman, M. (1996) Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *Journal of Molecular Evolution*, 43, 83–92.
Swofford, D.L. (2002) *PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.* (4 ed.). Sinauer Associates, Sunderland, Massachusetts.
Tang, K.L., Agnew, M.K., Chen, W.-J., Raley, M.E., Sado, T., Schneider, L.M. Yang, L., Bart, H. L., He, S., Liu, H., Miya, M., Saitoh, K., Simon, A. M., Wood, R. M. & Mayden, R. L. (2011) Phylogeny of the gudgeons (Teleostei: Cyprinidae: Gobioninae). *Molecular Phylogenetics and Evolution*, 61, 103–124.
Tao, W.J., Zou, M., Wang, X.Z., Gan, X.N., Mayden, R.L. & He, S.P. (2010) Phylogenomic analysis resolves the formerly intractable adaptive diversification of the endemic clade of East Asian Cyprinidae (Cypriniformes). *PLoS ONE*, 5 (10): e13508. http://dx.doi.org/10.1371/journal.pone.0013508.
Tsigenopoulos, C.S., Ráb, P., Naran, D. & Berrebi, P. (2002) Multiple origins of polyploidy in the phylogeny of southern African barbs (Cyprinidae) as inferred from mtDNA markers. *Heredity*, 88, 466–473.
Uyeno, T. & Smith, G.R. (1972) Tetraploid origin of the karyotype of catostomid fishes. *Science*, 175, 644–646.
Wilson, M.V.H. (1977) Middle Eocene freshwater fishes from British Columbia. *Royal Ontario Museum, Life Sciences Contributions*, 113, 1–61.
Yang, L., Arunachalam, M., Sado, T., Levin, B.A., Golubtsov, A.S., Freyhof, J., Friel, J.P., Chen, W.-J., Hirt, M.V., Manickam, R., Agnew, M.K., Simons, A.M., Saitoh, K., Miya, M., Mayden, R.L. & He, S. (2012) Molecular phylogeny of the cyprinid tribe Labeonini (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution*, 65, 362–379.
Zwickl, D.J. & Hillis, D.M. (2002) Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology*, 51, 588–598.