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

Richard L. Mayden [a,*], Wei-Jen Chen [a], Henry L. Bart [b], Michael H. Doosey [b], Andrew M. Simons [c], Kevin L. Tang [a], Robert M. Wood [a], Mary K. Agnew [a], Lei Yang [a], M. Vincent Hirt [d], Mark D. Clements [b,1], Kenji Saitoh [e], Tetsuya Sado [f], Masaki Miya [f], Mutsumi Nishida [g]

[a] *Department of Biology, Saint Louis University, 3507 Laclede Ave., St. Louis, MO 63103, USA*
[b] *Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA 70118, USA*
[c] *Department of Fisheries, Wildlife, and Conservation Biology and Bell Museum of Natural History, University of Minnesota, St. Paul, MN 55108, USA*
[d] *Ecology, Evolution and Behavior Graduate Program and Bell Museum of Natural History, University of Minnesota, St. Paul, MN 55108, USA*
[e] *Tohoku National Fisheries Research Institute, Fisheries Research Agency, 3-27-5 Shinhama-cho, Shiogama-shi, Miyagi 985-0001, Japan*
[f] *Department of Zoology, Natural History Museum and Institute, Chiba, 955-2 Aoba-cho, Chuo-ku, Chiba 260-8682, Japan*
[g] *Ocean Research Institute, University of Tokyo, 1-15-1 Minamidai, Nakano-ku, Tokyo 164-8639, Japan*

ARTICLE INFO

ABSTRACT

The order Cypriniformes is the most diverse clade of freshwater fishes and is natively distributed on all continents except South America, Australia, and Antarctica. Despite the diversity of the group and the fundamental importance of these species in both ecosystems and human culture, relatively little has been known about their relationships relative to their diversity. In recent years, with an international effort investigating the systematics of the group, more information as to their genealogical relationships has emerged and species discovery and their descriptions have increased. One of the more interesting aspects of this group has been a traditional lack of understanding of the relationships of the families, subfamilies, and other formally or informally identified groups. Historical studies have largely focused on smaller groups of species or genera. Because of the diversity of this group and previously published whole mitochondrial genome evidence for relationships of major clades in the order, this clade serves as an excellent group to investigate the congruence between relationships reconstructed for major clades with whole mitogenome data and those inferred from a series of nuclear gene sequences. As descent has resulted in only one tree of life, do the phylogenetic relationships of these major clades converge on similar topologies using the large number of available characters through this suite of nuclear genes and previously published mitochondrial genomes? In this study we examine the phylogenetic relationships of major clades of Cypriniformes using previously published mitogenomes and four putative single-copy nuclear genes of the same or closely related species. Combined nuclear gene sequences yielded 3810 bp, approximately 26% of the bp found in a single mitogenome; however homoplasy in the nuclear genes was measurably less than that observed in mitochondrial sequences. Relationships of taxa and major clades derived from analyses of nuclear and mitochondrial sequences were nearly identical and both received high support values. While some differences of individual gene trees did exist for species, it is predicted that these differences will be minimized with increased taxon sampling in future analyses.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

In theory, the process of descent has produced a single evolutionary history of life or tree of species (Wiley, 1981). Thus, one

should be able to reconstruct this history using any number of heritable character types (e.g., DNA, proteins, morphology, etc.) and appropriate phylogenetic methodologies (Mayden and Wiley, 1992). With the coalescence of advances in molecular biology, informatics, and computational biology, a new era of phylogenomics in systematic biology has emerged, wherein large numbers of DNA sequences originating from the nuclear, mitochondrial, and/ or plastid genomes can be available for the inference of the tree of life (Chen et al., 2004, 2008; Delsuc et al., 2005). While great ad-

vances have clearly been achieved in harvesting sequence data from one or more of these genomes for any given group of organisms, analyses of these types of data and their ultimate interpretations as evolutionary "indicators" of sister group relationships remain an area of much needed attention. Recent analyses of these heterogeneous genomic data do not always yield the theoretically predicted single evolutionary history. Instead, they can result in different species relationships, depending upon the gene(s) examined and are often referred to as gene trees.

Historically, genes from the mitochondrial (mtDNA) or plastid (pDNA) genomes formed the essence of most analyses in the early years of molecular systematics. The preference for these circular genomes over nuclear DNA (ncDNA) was largely due to their single copy nature, relative ease in sequencing, lack of complications associated with polyploidy and paralogy, relatively short generation time and time for coalescence in lineages, and a lack of recombination (Moore, 1995). Furthermore, even in the early years of this endeavor it was understood that evidence pointed to rates of anagenesis in nuclear exon regions as significantly slower than evolution of mtDNA and pDNA genes (Vawter and Brown, 1986), making the reconstruction of relationships of closely related species near impossible without these latter genes (Slade et al., 1994). Rates of evolution of intron regions of ncDNA could be comparable to mtDNA or pDNA but alignment of some of these potentially informative regions can be difficult, especially in situations involving comparisons of distantly related taxa (Chen et al., 2008). Alignments have been particularly problematic in some instances (Lutzoni et al., 2000; Kawakita et al., 2003) but less so in others (He et al., 2008a; Schönhuth et al., 2008).

For several reasons, nuclear genomes have historically been either too difficult or expensive for use in phylogeny reconstruction and evidence exists that their utility in achieving a tree of life may be of less importance than has been argued by some (Moore, 1995). Concerns with ncDNA have included the comparative use of orthologous versus paralogous genes and finding single-copy genes with rates of evolution appropriate for species-level phylogenies. Likewise, ncDNA presents potential difficulties in development of primers that are effective across a diverse array of taxa for both amplification and sequencing relative to some mtDNA genes for which some universal primers exist (Kocher et al., 1989; Palumbi, 1996; Folmer et al., 1994). However, for many, the maternal inheritance of mitochondrial or plastid genomes, their existence as a single linkage group (haplotype), and the possibility of hybridization, are all qualities that make these character sets suspect in the reconstruction of species relationships.

The issues surrounding the use of mtDNA, pDNA, or ncDNA are far from being resolved in systematic biology. Pros and cons of all these data types are evident in different studies addressing different types of questions under different circumstances. Even the argument of hybridization between species as a reason to restrict the use of mtDNA in favor of ncDNA is unreasonable given that both genomes are exchanged in hybridization events. However, those not favoring mtDNA and pDNA would likely argue that ncDNA is under a different selective regime than mtDNA, and that the former may be more likely represent the "true" tree of species with strong selection for function of nuclear loci and the latter could be retained as "silent" misinformation under limited selection in hybrid regions. Regardless, existing studies comparing these genomes generally entail comparisons of trees in closely related species where impacts of hybridization or lineage sorting from polymorphisms in ancestral species may be readily recognized by contradictory sister-species relationships. The correspondence of these differential sets of genes in reconstructing the evolutionary histories at higher levels of organization deserves more attention.

Species of the order Cypriniformes constitute the largest monophyletic group of freshwater fishes in the world, with at least 2600 species and an estimated 2500 species remaining to be described. Many of these species are economically important and critical food resources in countries where they serve as the primary protein source, other species serve as important research organisms (e.g., the zebrafish, *Danio rerio*), and all of the species are fundamental elements in the aquatic communities in lakes, rivers, streams, creeks, caves, and springs (Mayden et al., 2007, 2008). With the fundamental importance of these species in research, in natural and altered ecosystems, with human populations, and their widespread distributions across the globe, recent international efforts have begun to focus on their diversity and evolutionary relationships. Historically, this clade has been examined for single mtDNA gene phylogenies, usually cytochrome *b* (Schmidt et al., 1994; Dowling and Naylor, 1997; Briolay et al., 1998; Schmidt et al., 1998; Simons and Mayden, 1997, 1998, 1999; Zardoya and Doadrio, 1999; Broughton and Gold, 2000; Simons et al., 2000; Bielawski and Gold, 2001; Cunha et al., 2002; Machordom and Doadrio, 2001; Raley and Wood, 2001; Schönhuth et al., 2001; Durand et al., 2002; Schönhuth and Doadrio, 2003; Simons et al., 2003; He et al., 2004; Simons, 2004; Mayden et al., 2006; Tang et al., 2006; Schönhuth et al., 2007; Rüber et al., 2007). The recent interest in these fishes has resulted in phylogeny reconstructions using single mtDNA genes, whole mitochondrial genomes (Saitoh et al., 2003; Saitoh et al., 2006; Wang et al., 2007; He et al., 2008b; Miya et al., 2006) and most recently ncDNA sequences or limited combined analyses (Mayden et al., 2006, 2007; Schönhuth et al., 2008; Mayden et al., 2008; Šlechtová et al., 2008). No investigations have examined the relative merits in phylogeny reconstruction using both whole mitochondrial genomes in combination or separately with multiple ncDNA loci.

Herein, we examine these two general classes of genes and their relative abilities in reconstructing the relationships of the major clades within the Cypriniformes, and speculate as to the utility of individual mitochondrial genes, the whole genome, single-copy nuclear genes, and combined analyses for this diverse set of data. The analyses of this study are based on sequences of whole mitochondrial genomes published by Saitoh et al. (2006) and on sequences of exon regions from four putatively "single-copy" nuclear genes, including recombination activating gene 1 (RAG1), rhodopsin gene (RH), interphotoreceptor retinoid-binding protein gene 2 (IRBP2), and growth hormone gene (GH). Exon 3 of RAG1 and the protein-coding region from the intron-less rhodospin gene (for teleost taxa) are widely used nuclear markers in phylogenetic relationships for many groups of ray-finned fishes (e.g., López et al., 2004; Rüber et al., 2004; Holcroft, 2005; Sullivan et al., 2006; Zaragüeta-Bagils et al., 2002; Chen et al., 2003, 2007; Dettaï and Lecointre, 2004, 2005), including the Cypriniformes (Mayden et al., 2007, 2008; Schönhuth et al., 2008; Šlechtová et al., 2007, 2008). Exon 1 of IRBP2 is a recently developed nuclear locus useful for studying evolutionary relationships of the Cypriniformes (Chen et al., 2008) and the Acanthomorpha (spiny-rayed fishes) (Dettaï and Lecointre, 2008). Amino acid sequences from the protein-coding region of GH was first used for inferring the phylogeny of "bony" fishes by Bernardi et al. (1993) and subsequently within subfamilies for various groups of teleosts, including salmonines (Oakley and Phillips, 1999) and labeonines (Rajesh and Majumdar, 2007). Like salmonids, some cypriniforms (Catostomidae, certain groups of Cyprinidae) are tetraploids, putatively evolving from hybridization events early in the history of these two groups (Uyeno and Smith, 1972); however, this has yet to be tested in a phylogenetic context and thus the origin of the tetraploid nature of these fishes remains unknown. Recently published and ongoing work characterizing paralogous copies of GH in catostomids and some cyprinids demonstrates that genomic duplication in these two

groups are independent (Clements et al., 2004 this study). In this study we use newly developed primers for amplifying and sequencing most of the GH gene region (Exons 2–5) for a diverse array of cypriniform fishes. The primers designed and used herein permit us to unambiguously amplify and sequence a single orthologous copy of GH for catostomids and cyprinids.

This research is part of an ongoing international Tree of Life initiative on the order Cypriniformes, aimed at furthering our understanding of the evolution of life.

2. Materials and methods

2.1. Taxonomic sampling

Our taxonomic sampling attempted to match the previous whole mitogenomic study by Saitoh et al. (2006). In cases where the same species was not available, a congeneric representative (e.g., species from *Pangio*, *Xenocypris*, and *Phoxinus*) was chosen as a substitute, if possible. However, these substitute species were excluded from the combined analyses of nuclear data plus whole mitochondrial genome data. The datasets included between 46 and 49 cypriniform taxa depending on data set plus several representatives of other ostaryiophysans as outgroups (Table 1; including GenBank accession numbers of corresponding genes or genome sequences).

2.2. Laboratory molecular work

Tissue extraction was performed using Qiagen DNeasy extraction kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Extracted DNA quantity was measured using a spectrophotometer (Eppendorf). DNA amplification was conducted by PCR (Mullis and Faloona, 1987; Saiki et al., 1988) for fragments from the four targeted nuclear loci. Primers used in this study for PCR/sequencing RAG1, RH and IRBP2 have been published by Chen et al. (2003, 2007, 2008) and López et al. (2004). Methods for collecting DNA sequences from these three loci followed the procedures outlined in Chen et al. (2008). The laboratory protocol for collecting GH data is described as follows.

Two cycles of PCR were used to amplify GH from genomic DNA. The first round PCR was performed in a 15.5-µl reaction volume containing 9.725 µl sterile distilled water, 1.5 µl of 10× EX Taq buffer (Takara), 1.2 µl of dNTP (2.5 mM each), 1.0 µl each of primers GH22F and GHR (10 µM) (Clements et al., 2004), 0.075 µl EX Taq polymerase (Takara), and 1.0 µl of template containing about 25 ng DNA. The thermal cycle profile was as follows: (1) initial denaturation at 94 °C for 60 s; (2) then 30 cycles of denaturation at 94 °C for 30 s, annealing at 52 °C for 30 s, and extension at 72 °C for 120 s; and (3) final extension at 72 °C for 10 min. The first round PCR produced an amplicon ranging from 1200 to 2500 bp (depending on length of introns). First-round PCR products were diluted in an appropriate volume of purified water based on intensity of bands observed on an agarose gel and used as template for the short PCR. Short PCR was conducted using up to three internal primer pairs for the GH gene with the same reaction mixture. Thermal cycling profile was the same as first round PCR except extension time was reduced to 30 s. The 5′ untranslated region (UTR) and exons 1 through 3 were amplified and sequenced for catostomids using primers GHIF (Clements et al., 2004) and GH240R. The 5′ UTR and exons 1 through 3 were amplified and sequenced for cyprinids and cobitids with promoter primers GHp57F and GHp221F with GH264R, respectively. Thermal cycling profile was the same as first round PCR except extension time was reduced to 60 s. The amplification and sequencing primers and their presented oligonucleotides are listed in Table 2.

Short PCR products were purified with ExoSAP-IT (USB) and directly sequenced using Big Dye 3.1 terminator cycle sequencing kit (Applied Biosystems) and PCR primers. Sequence reaction products were purified by ethanol precipitation. Sequencing reactions were run and visualized on an ABI 3730XL automated DNA sequencer (Applied Biosystems).

2.3. Dataset and phylogenetic analysis

Aligned sequence data matrices for whole mitochondrial genome (excluding d-loop region) used in this study were obtained from co-authors, KS and MM. Data matrices used were exactly the same as those from Saitoh et al. (2006). Nuclear DNA sequences were edited and managed using Se-Al v2.0a11 (Rambaut, 1996). Compiled sequences were initially aligned with Clustal X (Thompson et al., 1997), then adjusted manually based on the inferred amino acid translation. Three operational datasets (1) using nuclear DNA sequences only; (2) using mitochondrial DNA sequences only; and (3) using DNA sequences of combined data from 1 and 2 were constructed for subsequent analyses. As phylogenetic analyses of protein-coding genes can be biased from homoplasy at third codon positions due to multiple substitutions in transitions (Saitoh et al., 2006) and/or due to base composition bias across taxa (Lockhart et al., 1994; Chen et al., 2003), we prepared two different types of character matrices with respect to character weighting schemes for data analyses. The first type of matrix included all targeted nucleotides from the compiled dataset without character weighting (or "equal-weighting scheme"). The second matrix included all targeted nucleotides from the compiled dataset, but with a "down-weighting scheme" used for the third codon position. This scheme (RY-coding) coded nucleotides A and G as a purine (R) and nucleotides T and C as a pyrimidine (Y).

Phylogenetic analyses were based on a partitioned Maximum Likelihood (ML) method and partitioned Bayesian approach (BA) for two different types of character matrices as implemented in the parallel version of RAxML (version 7.0.4) (Stamatakis, 2006) and MrBayes (version 3.1.1) (Huelsenbeck and Ronquist, 2001), respectively. The search of optimal ML trees and Bayesian analyses were performed on a high-performance cluster computing facility (with 32 nodes) located at SLU. We used mixed model analysis, which allows an individual model of nucleotide substitution to be estimated independently from each partition for the analyses. Partitions were assigned with respect to the codon positions of each nuclear protein-coding gene. Mitochondrial genome data were subdivided as five partitions according to Saitoh et al. (2006): first, second and third codon positions of protein-coding genes, ribosomal RNA genes (12S and 16S), and all tRNA genes. Likelihood ratio tests (Goldman, 1993), as implemented in MrModeltest 2.2 (Nylander, 2004), were used to choose models for each gene coding position in Partitioned BA.

The parameters for running MrBayes were set as follows: "lset nst = 6" (GTR), "lset nst = 2" (HKY), "lset nst = 1" (F81), "rates = invgamma" ($\Gamma + I$), or "rates = gamma" (Γ), "unlink" (unlinking of model parameters across data partitions), and "prset ratepr = variable" (rate multiplier variable across data partitions). Two independent Bayesian searches were conducted for each dataset. Four independent MCMC chains were performed with 3,000,000 replicates, sampling one tree per 100 replicates for each run. The distribution of log likelihood scores was examined to determine stationarity for each search and to decide if extra runs were required to achieve convergence in log likelihoods among runs or searches. We discarded initial trees with non-stationary log likelihood values as part of a burn-in procedure, and combined the remaining trees that resulted in convergent log likelihood scores from both independent searches. These trees were used to construct a 50% majority rule consensus tree. For the ML search

Table 1
Cypriniform taxa, genes, and Genbank accession numbers for nuclear gene sequences of representative species.

Family/subfamily	Taxon	GenBank Accession Numbers				
		Mt-genome	RAG1	RH	IRBP	GH
Outgroups						
Alestiidae	*Phenacogrammus interruptus*	AB054129	FJ197124	FJ197073	FJ197123	
Characidae	*Chalceus macrolepidotus*	AB054130	EU409607	EU409633	EU409665	
Chanidae	*Chanos chanos*	AB054133	AY430207	FJ197072		
Gonorynchidae	*Gonorynchus greyi*	AB054134	EU409606	EU409632	EU409664	
Callichthyidae	*Corydoras rabauti*	AB054128	FJ197125	FJ197074		
Bagridae	*Pseudobagrus tokiensis*	AB054127		FJ197075		
Clariidae	*Clarias batrachus*		DQ492521			AF416485
Heteropneustidae	*Heteropneustes fossilis*		DQ492522			AF416489
Cypriniformes						
Balitoridae	*Homaloptera leonardi*	AB242165	EU711130	FJ197027	FJ197076	FJ265022
Nemacheilidae	*Lefua echigonia*	AB054126	EF458305	FJ197028	FJ197077	FJ265023
	Schistura balteata	AB242172	EU711131	FJ197029	FJ197078	FJ265024
	Barbatula toni	AB242162	EU711133	FJ197030	FJ197079	
Vaillantellidae	*Vaillantella maassi*	AB242173	EU711132	FJ197031	FJ197080	FJ265058
Catostomidae	*Catostomus commersoni*	AB127394	EU409612	FJ197032	FJ197081	FJ265027
	Hypentelium nigricans	AB242169	EU711134	FJ197033	FJ197082	FJ265055
	Minytrema melanops	AB242166	EU711135	FJ197034	FJ197083	FJ265050
	Cycleptus elongatus	AB126082	EU409613	FJ197035	FJ197084	FJ265028
	Myxocyprinus asiaticus	AP006764	EU711136	FJ197036	FJ197085	FJ265052
Botiidae	*Chromobotia macracantha*	AB242163	EU711137	FJ197037	FJ197086	
	Leptobotia mantschurica	AB242170	EU711138	FJ197038	FJ197087	FJ265035
Cobitidae	*Acantopsis choirorhynchos*	AB242161	EU711139	FJ197039	FJ197088	
	Cobitis striata (large race)	AB054125	Saitoh unpublished	Saitoh unpublished	Saitoh unpublished	
	Misgurnus nikolskyi	AB242171	EU711140	FJ197040	FJ197089	FJ265057
	Pangio anguillaris	AB242168			FJ197090	
	Pangio oblonga		EU711141	FJ197041	FJ197091	
Cyprinidae						
Acheilognathinae	*Acheilognathus typus*	AB239602	EU292688	FJ197042	FJ197092	FJ265056
Acheilognathinae	*Rhodeus ocellatus kurumeus*	AB070205	EU711142	FJ197043	FJ197093	FJ265041
Leuciscinae	*Alburnus alburnus*	AB239593	EU711143	FJ197044	FJ197094	FJ265045
Leuciscinae	*Pelecus cultratus*	AB239597	EU711144	FJ197045	FJ197095	FJ265051
Cultrinae	*Ischikauia steenackeri*	AB239601	EU292687	FJ197046	FJ197096	FJ265033
Cultrinae	*Chanodichthys mongolicus*	AP009060	EU711145	FJ197047	FJ197097	FJ265046
Cyprininae	*Barbonymus gonionotus*	AB238966	EU711146	FJ197048	FJ197098	FJ265053
Cyprininae	*Barbus barbus*	AB238965	EU711147	FJ197049	FJ197099	
Cyprininae	*Barbus trimaculatus*	AB239600	EU711148	FJ197050	FJ197100	FJ265059
Cyprininae	*Cyprinus carpio*	AP009047	AY787040	U02475/Z71999	FJ197101	FJ265047
Cyprininae	*Gymnocypris przewalskii*	AB239595	EU711149	FJ197051	FJ197102	
Cyprininae	*Labeo batesii*	AB238967	EU711150	FJ197052	FJ197103	
Cyprininae	*Labeo senegalensis*	AB238968	EU711151	FJ197053	FJ197104	FJ265034
Cyprininae	*Puntius ticto*	AB238969	EU711152	FJ197054	FJ197105	
Cyprininae	*Carassius auratus*	AB006953	DQ196520/DQ196518	L11863	X80802	AF069398
Gobioninae	*Gnathopogon elongatus*	AB218687	EU711153	FJ197055	FJ197106	FJ265030
Gobioninae	*Gobio gobio*	AB239596	EU292689	FJ197056	FJ197107	FJ265048
Gobioninae	*Hemibarbus barbus*	AB070241	EU711154	FJ197057	FJ197108	FJ265032
Gobioninae	*Pseudorasbora pumila pumila*	AB239599	EU711155	FJ197058	FJ197109	FJ265039
Gobioninae	*Pungtungia herzi*	AB239598	EU711156	FJ197059	FJ197110	FJ265040
Gobioninae	*Sarcocheilichthys variegatus microoculus*	AB054124	EU711157	FJ197060	FJ197111	FJ265042
Leuciscinae	*Cyprinella lutrensis*	AB070206	EU711158	FJ197061	FJ197112	FJ265061
Leuciscinae	*Notemigonus crysoleucas*	AB127393	EF452831	FJ197062	FJ197113	FJ265036
Leuciscinae	*Tribolodon nakamurai*	AB218896	EU711159	FJ197063	FJ197114	FJ265044
Xenocyprinae	*Xenocypris argentea*	AP009059				
Xenocyprinae	*Xenocypris macrolepis*		EU711160	FJ197064	FJ197115	FJ265049
Leuciscinae	*Phoxinus phoxinus*		EU711161	FJ197065	FJ197116	FJ265038
Leuciscinae	*Phoxinus perenurus*	AP009061				
Rasborinae	*Aphyocypris chinensis*	AB218688	EU292692	FJ197066	FJ197117	FJ265026
Rasborinae	*Esomus metallicus*	AB239594	EU292702	FJ197067	FJ197118	FJ265029
Rasborinae	*Opsariichthys uncirostris*	AB218897	FJ197126	FJ197068	FJ197119	FJ265037
Rasborinae	*Zacco sieboldii*	AB218898	EU292713	FJ197069	FJ197120	FJ265025
Rasborinae	*Danio rerio*	AC024175	U71093	L11014	X85957	AJ937858
Tincinae	*Tinca tinca*	AB218686	EU711162	FJ197070	FJ197121	FJ265043
Gyrinocheilidae	*Gyrinocheilus aymonieri*	AB242164	EU292682	FJ197071	FJ197122	FJ265031

with the mixed model of nucleotide substitution, we used the GTR + Γ + I model (with four discrete rate categories) for each partition because RAxML only provides GTR related models (GTR + Γ, GTR + Γ + I and GTR + CAT approximation) of rate heterogeneity for nucleotide data (Stamatakis, 2006). The ML tree search was conducted by performing 100 distinct runs using the default algo-

Table 2
Growth hormone (GH) PCR primers for amplification and sequencing of cyprinifomres and expected length of amplicons.

Name	Sequence (from 5′ to 3′)	Expected length (bp)
First round PCR primers		
GH22F	YTGTCKDTGGTSCTGGTYAGT	
GHR[a]	CAGGGTRCAGTTKGAATCSAR	1200–2500
Short PCR primers		
GH22F	YTGTCKDTGGTSCTGGTYAGT	
GH264R	GCTYTYTYTGBGTTTCATSTTT	500–800
GH181F	CAGCTGAGTAAAATCTTYCCT	
GH295R	CTCCCARGAYTCAATGAGGYG	400–1400
GH274F	AAGCTBCTTCGYATCTCYTT	
GHR[a]	CAGGGTRCAGTTKGAATCSAR	500–800
5′ UTR PCR primers		
GHIF[a]	AAAGCCTTCAACTAAGACTAAC	
GH240R	TTCTGGGTTTCATGTTTGTCA	700–750
GHp57F	CAAACACTCACAAGCCTTCAAC	600–1050
GHp221F	CAAATACATGACCGGAACTAC	1100–1350

[a] From Clements et al. (2004).

rithm of the program from random trees (-d option) as a starting tree for each run. The final tree was determined by a comparison of likelihood scores under the GTR + Γ + I model among suboptimal trees obtained per run.

Nodal support was assessed using the bootstrap (BS) procedure (Felsenstein, 1985) under Maximum Parsimony (MP) and Maximum Likelihood (ML) criterion, based on 1000 pseudo-replicates and the resulting *a posteriori* probabilities from partitioned BA. The studies on experimental simulation suggest that, being more conservative, the non-parametric bootstrap approach might be less prone to strongly supporting a false phylogenetic hypothesis, while posterior probabilities put overconfidence on a given phylogenetic hypothesis (Douady et al., 2003). In the present study, we set up posterior probabilities and non-parametric bootstrap support (especially from MPBS) as potential upper and lower bounds of node robustness for our inferred phylogenetic trees. For the MPBS analyses using PAUP∗, optimal trees were obtained by heuristic searches with random stepwise addition sequences followed by TBR swapping for 100 replications (Swofford, 2002). The MLBS results (through analyses using RAxML web-servers) (Stamatakis et al., 2008) were obtained from the CIPRES cluster (CIPRES Portal v 1.13) at the San Diego Supercomputer Center at http://www.phylo.org/sub_sections/portal/. The retention index and chi-square tests of base composition stationarity were calculated using PAUP∗.

Finally, to gain further insight on 'phylogenetic' signals presented in each data set across taxa, a protocol called repeated-bootstrap components (Chen et al., 2003) was employed. This method consisted of scoring the repeated clades found in the listing of bootstrap bipartitions from MP analysis produced by PAUP∗ for each separate gene partition (repeated bootstrap components) using the computer program developed by the author (W.J.C.). Resulting repeated bootstrap components were mapped onto the optimal ML tree derived from simultaneous analysis of the combined dataset with equal-weighting scheme (Fig. 3, left). The corresponding bootstrap values for each data partition were displayed in the form of a histogram for each node on the tree (Fig. 4).

3. Results

A total of 3810 bp were aligned for the exon regions of four nuclear genes for 56 taxa (including seven outgroups) sampled in this study. The length of aligned sequences from each locus was 1497 bp (RAG1), 831 bp (RH), 849 bp (IRBP), and 633 bp (GH). No internal indels were found among aligned sequences of ingroup

taxa. Of these, 1826 sites were constant and 1618 sites were parsimony informative. The nuclear dataset is about 26% the size of the whole mt-genome dataset (14,563 bp) of Saitoh et al. (2006). The level of homoplasy contained in the nuclear dataset, as measured by RI, was much lower than that observed in the mitochondrial dataset (RI = 0.61 vs 0.33 for nuclear and mitochondrial datasets, respectively). An excess of nucleotide substitutions at third codon positions of the mitochondrial protein-coding genes is likely one of the causes for the higher level of homoplasy in the mitochondrial dataset (Saitoh et al., 2006; Chen et al., 2008). Additionally, when base composition varies significantly among taxa, all classical methods of phylogenetic construction tend to group sequences of similar nucleotide composition together, regardless of evolutionary history (Lockhart et al., 1994; Chen et al., 2003). Base composition stationarity could not be rejected on all sites for each codon position of nuclear genes except for the third codon position of RAG1 and rhodopsin. Thus, results obtained from the RY-coding analyses were compared to those from equal-weighting analyses to assess the consistency of phylogenetic inferences.

Relationships of taxa derived from partitioned ML and Bayesian analyses of nuclear and mitochondrial sequences were nearly identical; only likelihood trees are presented herein (Figs. 1–3). Analyses of individual nuclear genes yielded different tree topologies (see Supplementary Material); however, the major clades identified among cypriniform species and sister group relationships of all the trees were consistent (Table 3). Several previously recognized groups such as cypriniform families and subfamilies (except for Rasborinae) were resolved as monophyletic in separate gene and/or genome trees. In addition, some clades were discovered that suggest relationships that differ from more traditional perspectives on cypriniform relationships. These include a clade (Clade V) grouping *Xenocypris* and cultrines, a clade (Clade VI) grouping some *Aphyocypris*, *Opsariichthys*, and *Zacco* plus members from the previous clade, and a clade (Clade VII) grouping *Tinca*, gobioninies, leuciscines and acheilognathines. The latter clade was not found in mt-genome trees but is recurrently found by two nuclear genes (RAG1 and GH) and in the phylogeny based on combined sequences of all nuclear genes (Table 3).

Separate phylogenetic signals presenting in each data set across taxa were assessed using the approach of repeated bootstrap components (Fig. 4). Given that bootstrap values are regarded as generally good measures of hierarchical signal (Hillis and Bull, 1993) under the criterion of repeatability (Chen et al., 2003), the histogram mapped on the tree for each node can be interpreted as the contribution to phylogenetic signal of each data partition in support of a corresponding node (Fig. 4). This information may also be useful for future studies of particular cypriniform groups by focusing on signal-rich genes when the target taxonomic samples become available. In general, given the results shown in the Fig. 4 the phylogenetic signal is not always homogeneously distributed across different data partitions throughout the tree. In general, MT data provided more pronounced signals in resolving terminal clades relative nuclear genes. In addition, the phylogenetic signal contribution at this taxonomic level for GH was rather weak in some case while that of RAG1 was strong. Regarding particular regions of the tree, RH appeared to perform relatively poorly for recovering several clades within the Cyprinidae, yet performed very well in recovering some clades within the loach group. Nonetheless, these analytical results were corroborated with the repeated clades identified from the individual gene/genome analyses (Table 3). With this approach we confirmed the monophyly of clades II and III where MT data provided no or contradictory signals relative to the overall signals from independent nuclear gene loci. Accordingly, the position of a critical taxa, *Vaillantella*, was resolved as the sister-group of cobitids, balitorids plus nemacheilids.
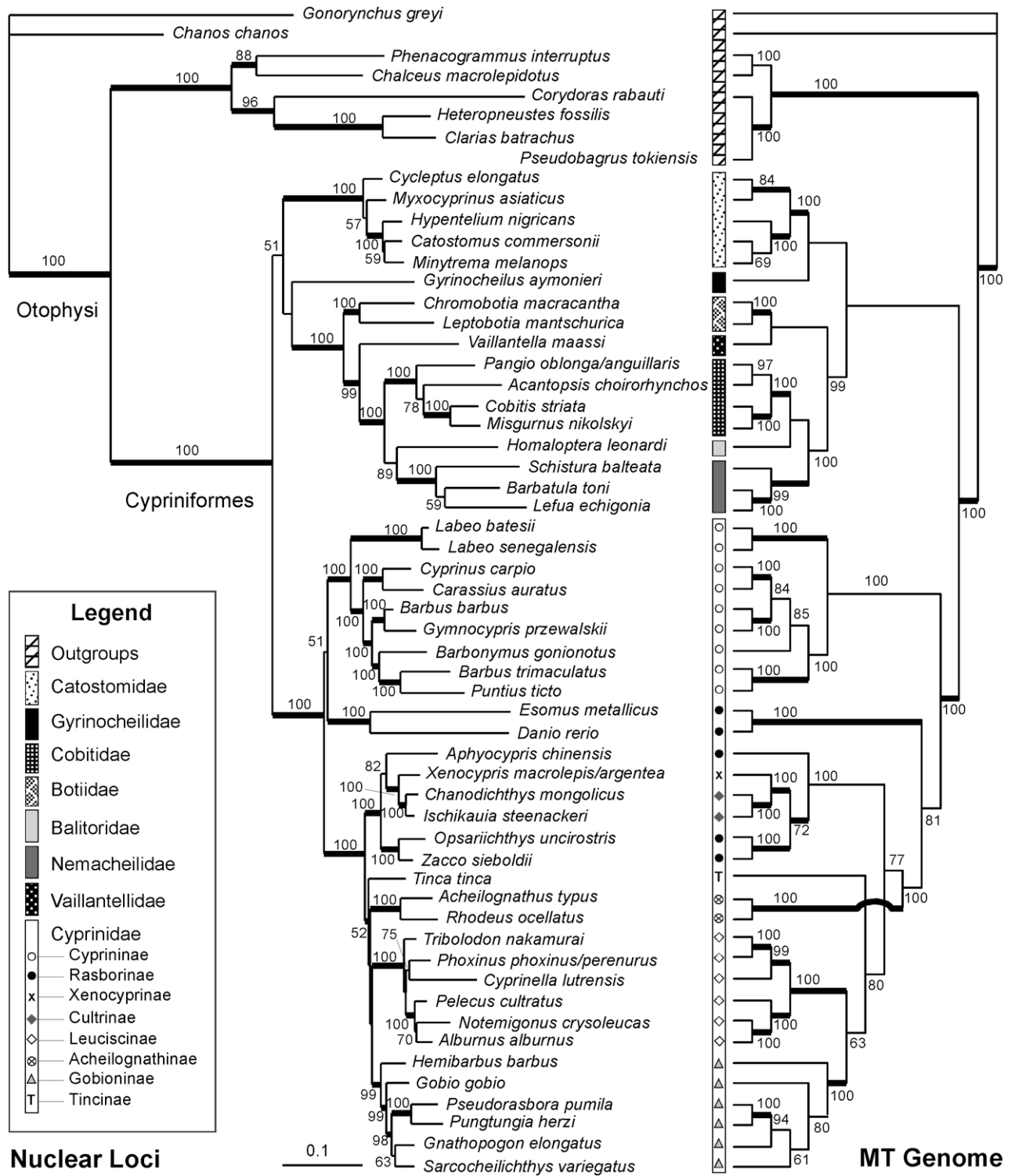
Fig. 1. Phylogenetic trees depicting relationships of the major clades within the Cypriniformes obtained from partitioned ML analyses and equal weighting of operational dataset 1 (all nc-DNA sequences; ca., 3810 bp) (left) and operational dataset 2 (mt-DNA sequences; ca.,14,563 bp) (right). ML scores of the trees are −45518.194337 and −284837.240795, respectively. Branch lengths are proportional to inferred character substitutions under the GTR + Γ + I model. Numbers on branches are ML bootstraps. Values below 50% are not shown. Bold branches on topologies indicate statistically robust nodes with *a posteriori* probabilities from partitioned Bayesian analysis ⩾0.95 and resulting MP bootstraps ⩾ than 80%.

All of the combined and separate nuclear gene analyses resulted in similar high bootstrap (from MLBS and/or MPBS) and posterior probability support for the monophyly of the order Cypriniformes and the families Catostomidae, Cobitidae, Botiidae, Nemacheilidae, and Cyprinidae; Gyrinocheilidae, Balitoridae, and Vaillantellidae were represented only by *Gyrinocheilus*, *Homaloptera*, and *Vaillantella*, respectively (Table 3, Figs. 1 and 2). Within Cyprinidae, the traditionally recognized subfamilies Acheilognathinae, Cyprininae, Cultrinae, Gobioninae, and Leuciscinae, were always resolved as monophyletic with strong nodal support;
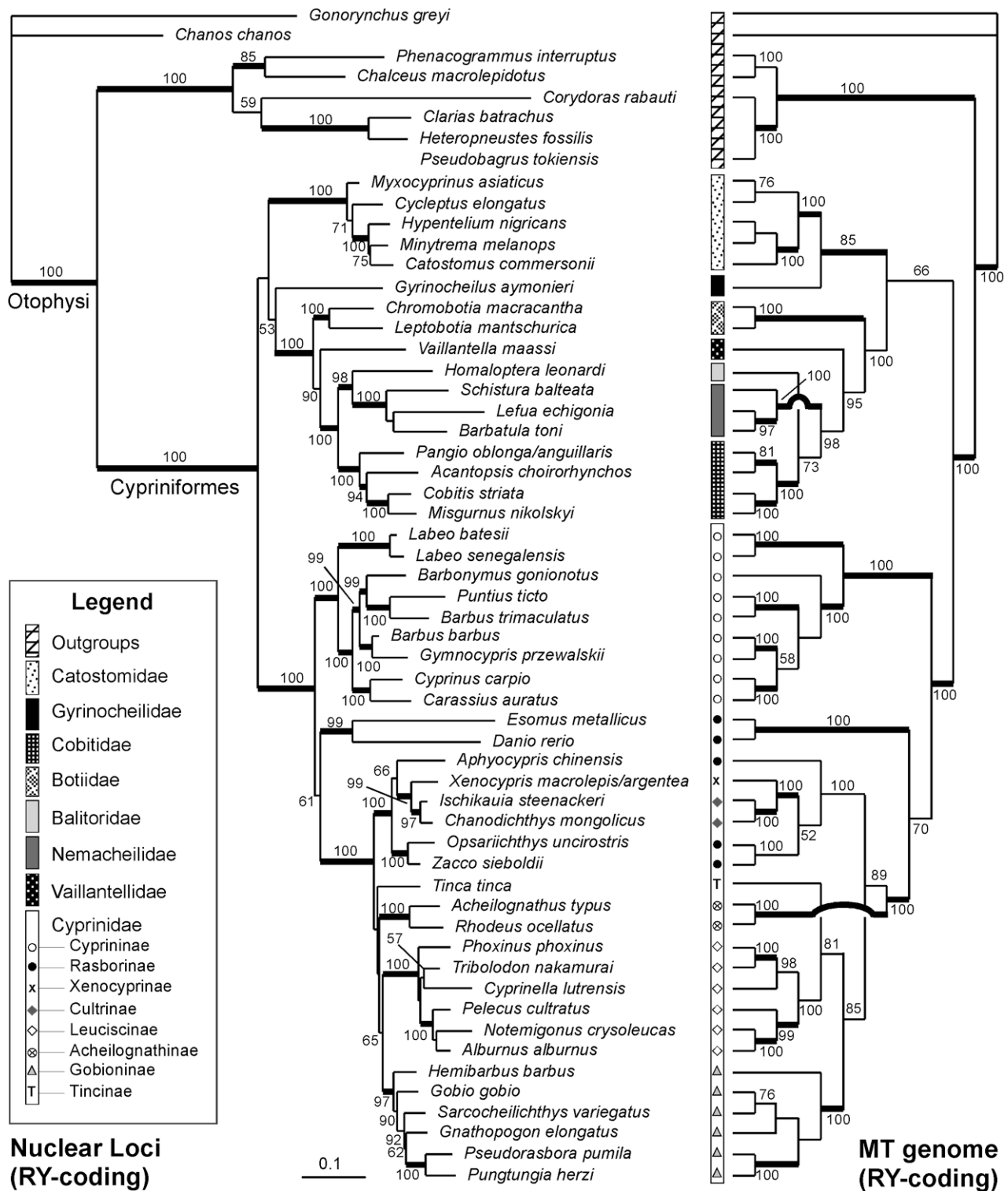
Fig. 2. Phylogenetic trees depicting relationships of the major clades within the Cypriniformes obtained from partitioned ML analysis and RY-coding of operational dataset 1 (all of nc-DNA sequences; ca., 3810 bp) (left) and operational dataset 2 (mt-DNA sequences; ca.,14,563 bp) (right). ML scores of the trees are −25994.707827 and −193917.795931, respectively. Branch lengths are proportional to inferred character substitutions under the GTR + Γ + I model. Numbers on branches are ML bootstraps. Values below 50% are not shown. Bold branches on topologies indicate statistically robust nodes with *a posteriori* probabilities from partitioned Bayesian analysis ⩾0.95 and resulting MP bootstraps ⩾ than 80%.

Tincinae and Xenocyprinae were only represented by *Tinca* and *Xenocypris*, respectively. The subfamily Rasborinae was never recognized as a monophyletic group and the included species were always resolved as three different, unrelated clades or lineages. Relationships of supraspecific groupings within the order, sup-

ported by combined nuclear gene variation, were similar for equal-weighting and RY-coding analyses; the superfamilies Cobitoidea and Cyprinoidea were resolved as reciprocally monophyletic groups. Cyprinoidea (or Cyprinidae) was strongly supported. Nodal support for Cobitoidea was poor; this clade
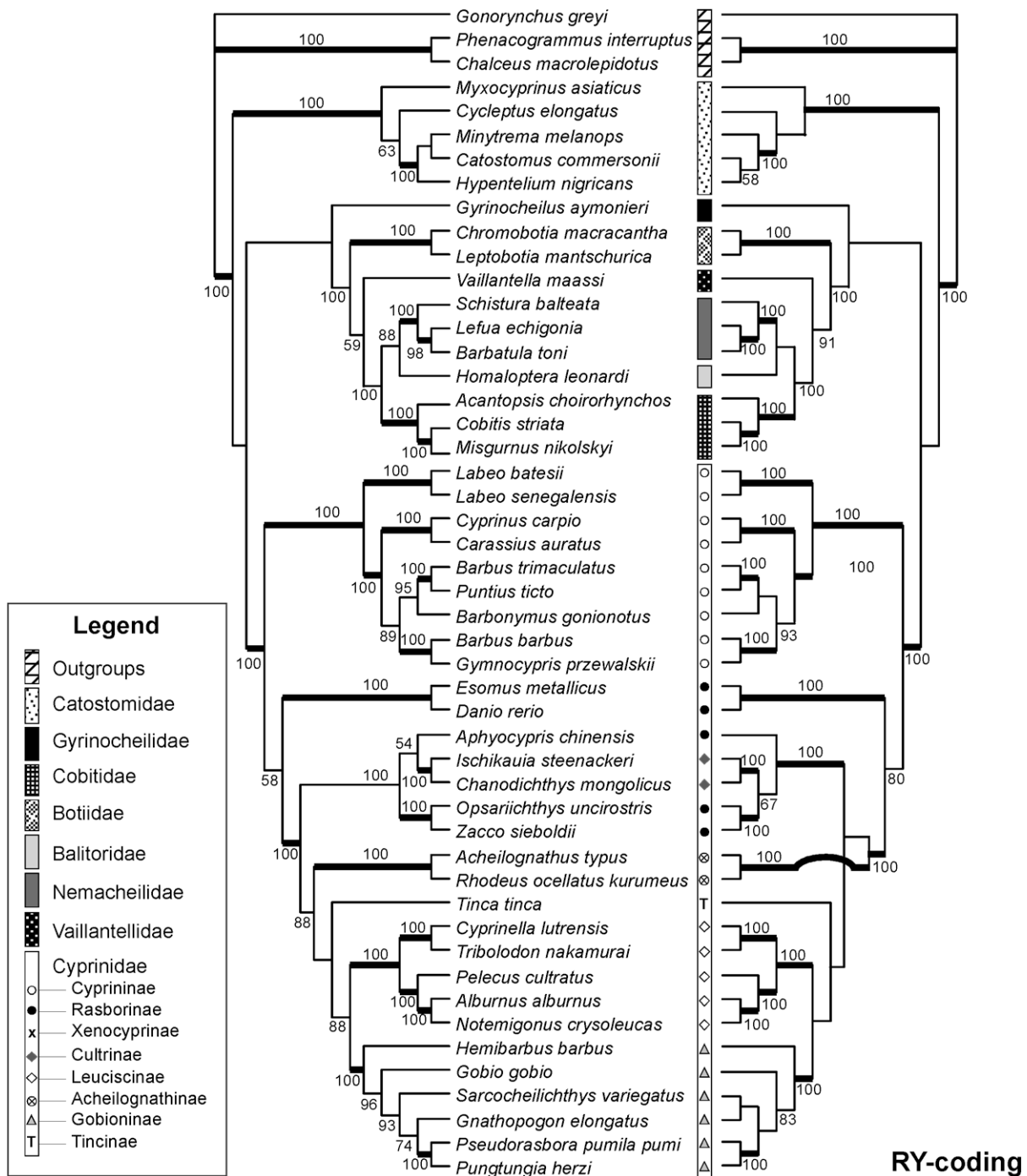
Fig. 3. Phylogenetic trees depicting relationships of the major clades within the Cypriniformes obtained from partitioned ML analysis and equal-weighting (left) and RY-coding (right) of operational dataset 3 (combined nc and mt-DNA sequences; ca., 18,354 bp). ML scores of the trees are −296598.143911 and −197655.988408, respectively. Numbers on branches are ML bootstraps. Values below 50% are not shown. Bold branches on topologies indicate statistically robust nodes with *a posteriori* probabilities from partitioned Bayesian analysis ⩾0.95 and resulting MP bootstraps ⩾ than 80%.

was recovered in two out of four nuclear gene analyses and mt genome analysis (Table 3). In the combined nc-mtDNA analysis, however, Cobitoidea was paraphyletic with respect to Cyprinoidea; Catostomidae was the sister group to all other cypriniforms (Fig. 3). Repeatability of clades recovered in independent gene trees has been argued as a criterion of support for particular topologies recovered in phylogenetic analyses (Chen et al., 2003). Based on this criterion, Cobitoidea is herein tentatively recognized as monophyletic, irrespective of the weak nodal support in ncDNA analysis and the combined nc-mtDNA analysis.

Within Cobitoidea, Gyrinocheilidae formed the sister group to the clade formed by Botiidae, Cobitidae, Vaillantellidae, Nemacheilidae and Balitoridae in combined nuclear gene trees, RAG1 gene trees and rhodopsin gene tree (with RY-coding analysis only), but this relationship received no bootstrap support and was not recovered by mtDNA analyses (Figs. 1 and 2; Table 3). In contrast to these findings from ncDNA analyses, Gyrinocheilidae was identified as the sister group to Catostomidae, but with relatively poor bootstrap support, in both equal-weighting and RY-coding analyses of mtDNA datasets. Botiidae likewise formed the sister group
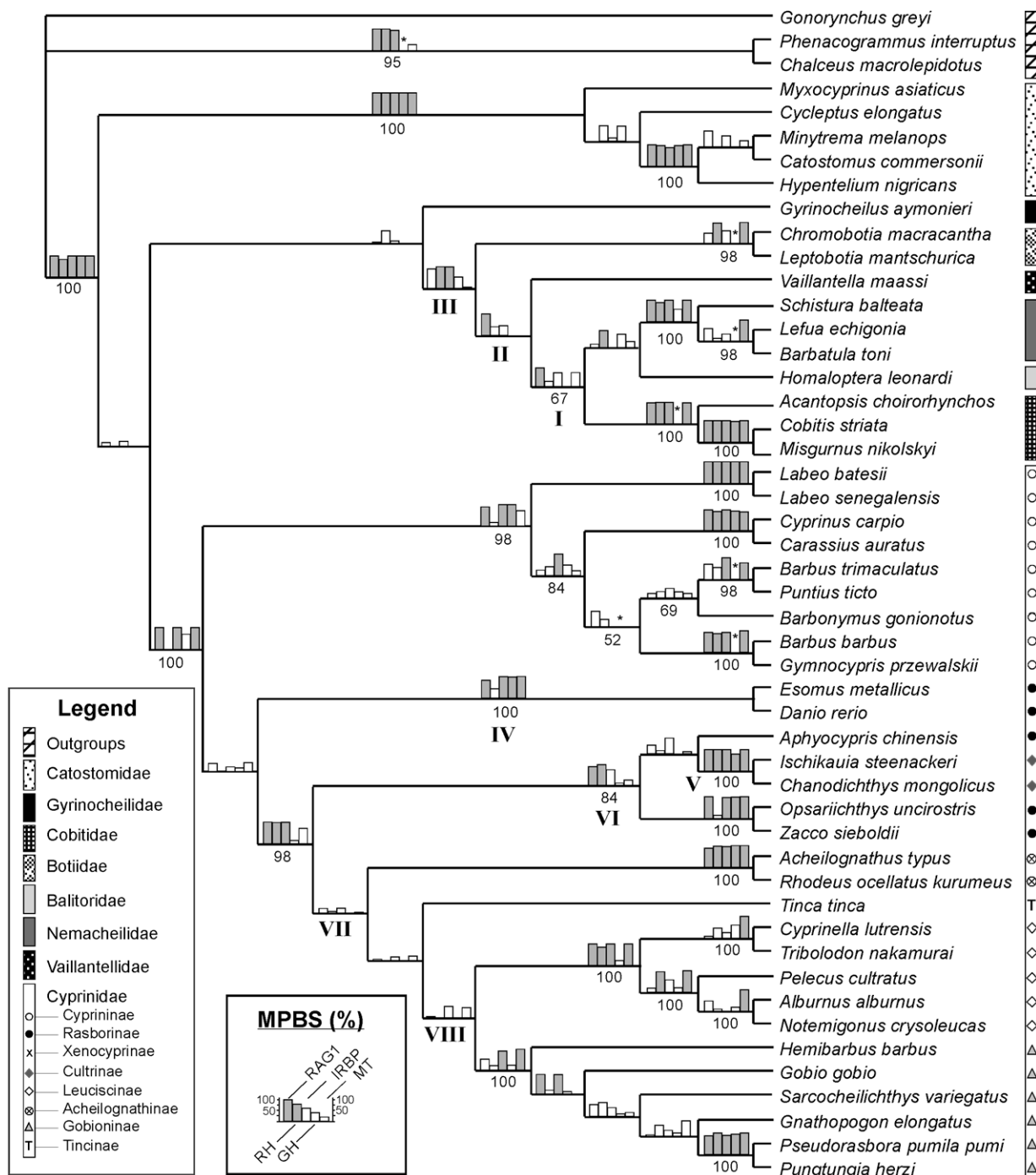
Fig. 4. General assessment of phylogenetic signals for each nuclear gene and mitochondrial genome at nodes across all taxa based on the repeated bootstrap components approach (see Section 2). Resulting bootstrap components (small histograms over branches) from the analyses of each gene/genome locus were mapped onto ML tree of combined, equal-weighting analysis as presented in Fig. 3 (left). Gray histograms indicate MP bootstraps ≥ than 75%. ∗ indicates values are not available because GH sequence data for the corresponding nodes are partially missing. Numbers below branches are resulting MPBS from the combined analysis. Values below 50% are not shown.

to Vaillantellidae plus the clade of Cobitidae, Nemacheilidae and Balitoridae in most of the analyses with generally robust support except for the equal-weighting analysis from mtDNA. The sister group relationship between Vaillantellidae and Cobitidae, Nemacheilidae plus Balitoridae received 99% ML bootstrap support in equal-weighting nuclear analyses, but only 90% support under the RY-coding scheme. The same sister group relationship was only found in the RY-coding analysis of mtDNA with 95% of ML bootstrap support. The sister relationships between Cobitidae and

Nemacheilidae plus Balitoridae received the highest bootstrap support in all analyses. However, Balitoridae was not sister to Nemacheilidae and was instead the sister group to Cobitidae in all mtDNA analyses, except in the combined analysis with ncDNA under both the equal-weighting and RY-coding schemes. All nuclear gene analyses identified Balitoridae as the sister group to Nemacheilidae (89–98%). The monophyly of Balitoridae plus Nemacheilidae was repeatedly determined in separate gene tree analyses (Table 3, Fig. 4).

Table 3
Summary results of major clades among cypriniforms identified from separate gene/genome trees inferred using partitioned ML analysis. Below each gene or genome support for the relationships on the left is indicated; a bold Y indicates that the resulting Maximum Likelihood bootstrap for the group is >80%; Y or N within parentheses denotes the result based on the same analysis but with RY-coding.

Hypothesis	Gene or genome					
	RAG1	RH	IRBP	GH	Nc	MT
Monophyly of						
Cypriniformes	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)
Cobitoidea	Y(Y)	**Y**(**Y**)	N(N)	N(N)	Y(Y)	Y(Y)
Cyprinoidea (Cyprinidae)	Y(Y)	N(N)	**Y**(**Y**)	Y(Y)	**Y**(**Y**)	**Y**(**Y**)
Catostomidae	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)
Cobitidae	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)
Botiidae	**Y**(**Y**)	**Y**(**Y**)	Y(Y)	na	**Y**(**Y**)	**Y**(**Y**)
Balitoridae	na	na	na	na	na	na
Nemacheilidae	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	Y(N)	**Y**(**Y**)	**Y**(**Y**)
Balitoridae + Nemacheilidae	N(Y)	**Y**(**Y**)	N(Y)	Y(N)	**Y**(**Y**)	N(N)
Clade I (Balitoridae, Nemacheilidae + Cobitidae)	**Y**(**Y**)	N(N)	**Y**(**Y**)	Y(Y)	**Y**(**Y**)	**Y**(**Y**)
Clade II (*Vaillantella* + Clade I)	**Y**(**Y**)	Y(N)	N(N)	N(N)	**Y**(**Y**)	N(**Y**)
Clade III (Botiidae + Clade II)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)
Acheilognathinae	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)
Cyprininae	**Y**(**Y**)	Y(Y)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)
Cultrinae	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	Y(Y)	**Y**(**Y**)	**Y**(**Y**)
Gobioninae	**Y**(**Y**)	N(N)	**Y**(**Y**)	N(N)	**Y**(**Y**)	**Y**(**Y**)
Leuciscinae	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	Y(Y)	**Y**(**Y**)	**Y**(**Y**)
Rasborinae	N(N)	N(N)	N(N)	N(N)	N(N)	N(N)
Clade IV (*Danio* + *Esomus*)	Y(**Y**)	**Y**(**Y**)	**Y**(N)	Y(Y)	**Y**(**Y**)	**Y**(**Y**)
Clade V (*Xenocypris* + Cultrinae)	**Y**(**Y**)	**Y**(**Y**)	**Y**(**Y**)	Y(Y)	**Y**(**Y**)	**Y**(**Y**)
Clade VI (*Aphyocypris*, *Opsariichthys*, *Zacco*, + Clade V)	**Y**(**Y**)	**Y**(**Y**)	**Y**(Y)	N(Y)	**Y**(**Y**)	**Y**(**Y**)
Clade VII (*Tinca*, Gobioninae, Leuciscinae and Acheilognathinae)	Y(**Y**)	N(N)	N(N)	Y(Y)	Y(Y)	N(N)
Clade VIII (Gobioninae + Leuciscinae)	N(N)	N(N)	Y(Y)	N(N)	Y(Y)	Y(N)
Acheilognathinae sister to						
Tinca + Clade VI + Clade VIII	N(N)	N(N)	N(N)	N(N)	N(N)	**Y**(**Y**)
Clade IV sister to						
Clade VI + VII	**Y**(Y)	N(N)	Y(N)	Y(Y)	N(Y)	Y(Y)
Gyrinocheilus sister to						
Catostomidae	N(N)	N(N)	N(Y)	N(N)	N(N)	**Y**(**Y**)
Gyrinocheilus sister to						
Clade III	Y(Y)	N(Y)	N(N)	N(N)	Y(Y)	N(N)
Vaillantella groups with						
Balitoridae, Nemacheilidae and Cobitidae	**Y**(**Y**)	Y(N)	N(N)	N(N)	**Y**(**Y**)	N(**Y**)
Vaillantella sister to						
Botiidae	N(N)	N(N)	Y(Y)	N(N)	N(N)	Y(Y)

Abbreviations used: RAG1, recombination-activating gene 1; RH, rhodopsin gene; IRBP, interphotoreceptor retinoid-binding protein gene; GH, growth hormone gene; Nc, nuclear genome; MT, mitochondrial genome; Y, Yes; N, No; na, not applicable.

Within Cyprinidae (or Cyprinoidea), the subfamily Cyprininae was always resolved as monophyletic. ML support for the monophyly of Cyprininae was 100% both in ncDNA and mtDNA analyses (Figs. 1 and 2) and in the combined nc-mtDNA analysis (Fig. 3). Cyprininae formed the basal sister group to other cyprinids in all analyses except in the equal-weight analysis of ncDNA (Figs. 1–3). The subfamily Rasborinae was never resolved as monophyletic and representatives were part of three separate clades or lineages (Figs. 1–3). The strongly supported clade inclusive of *Danio* and *Esomus* formed the sister group to Cyprininae in the equal-weight analysis of ncDNA (Fig. 1), and the sister group to the remaining cyprinids (excluding the Cyprininae) in the RY-coding analysis of ncDNA (Fig. 2), both coding analyses of mtDNA (Figs. 1 and 2), and of the combined nc-mtDNA (Fig. 3). Regardless of the analysis, relationships of the clade inclusive of *Danio* and *Esomus* received poor support in our analyses, but the sister group relationship between this clade and a clade grouping remaining cyprinids (without cyprinines) was confirmed by repeated clade analyses of nuclear genes (Table 3). The remaining cyprinid subfamilies, the clade inclusive of Cultrinae and Xenocyprinae (Clade V), and the clade (Clade VI) aligning Clade V with three remaining "rasborines" (*Aphyocypris* and *Opsariichthys*, *Zacco*) was generally resolved as a monophyletic group with strong nodal support (Table 3) (Supple-

mentary Material) (Figs. 1 and 2). Within Clade VI, *Opsariichthys* and *Zacco* were sister taxa. *Aphyocypris* was sister to Clade V in both the ncDNA analysis and the combined nc-mtDNA analysis with equal-weighting scheme. The monophyly of a clade (Clade VII) grouping Acheilognathinae, Tincinae, Leuciscinae, and Gobioninae received poor nodal support but this clade was confirmed in results of separate gene/genome analyses (Table 3). However, the relationships of deep branches within this clade remained unresolved.

In general, support for the monophyly of clades using only mtDNA varied but was quite high (Figs. 1–3). However, the number of nodes supported by high ML and/or MP bootstrap (equal to 80% and higher) and by high a posteriori probabilities (equal to 95% and higher) are relatively few in comparison with results from ncDNA analyses with a smaller data set. Regardless, relationships among clades as determined by mtDNA sequence variation was generally similar to those derived using ncDNA, except in the following instances and others already mentioned above. The basal lineage of Catostomidae included *Myxocyprinus* plus *Cycleptus* in mtDNA analyses (76–84%); however, *Myxocyprinus* formed the sister group to *Cycleptus* plus other catostomids in all combined nc-mt analyses but with very low support. Within the Cyprininae, *Labeo* (with two taxa represented) always formed a monophyletic group and was

sister to the remaining cyprinines, and *Barbonymus* varied in its placement depending on the genes and method of analysis. *Barbus* never formed a monophyletic group, with *Barbus barbus* more closely related to *Gymnocypris* and *Barbus trimaculatus* more closely related to *Puntius ticto*.

Contradictions among datasets in relationships of supraspecific groupings within family Catostomidae and subfamily Cyprininae could potentially result from the polyploid nature of the genomes in these taxa (Chen et al., 2008). In fact, a ncDNA tree could be inferred from a dataset with a mixture of orthologous and paralogous copies, while a mtDNA tree can be referred to an organismal history from maternal lines only. Both of these clades contain species (suckers, carps and goldfishes) with polyploid genomes (Ohno et al., 1967; Yu et al., 1987; Larhammar and Risinger, 1994; David et al., 2003). Thus, interpretation of some of the phylogenetic results reported herein should be treated with caution. Ultimately, unambiguous identification for paralogous loci of genes will be required for future studies to resolve the organismal phylogeny and to assess the evolutionary origin of polyploidy events.

4. Discussion

The relationships of major clades within the Cypriniformes were first examined using molecular characters by Saitoh et al. (2006) using whole mitochondrial genomes for 50+ ingroup and outgroup taxa. That paper and a paper published by Šlechtová et al. (2007) suggested novel relationships among cypriniform fishes requiring recognition of new families and indicating other unresolved taxonomic issues. Herein we recognize eight families within Cypriniformes (Šlechtová et al., 2007; Conway et al., in press). These include: Cyprinidae, containing the carps and minnows; Catostomidae, containing the suckers; Gyrinocheilidae, containing the algae eaters; Botiidae, containing the botiid loaches (formerly included in Cobitidae); Cobitidae, containing the cobitid loaches; Balitoridae, containing the flat loaches; Nemacheilidae, containing the nemacheilid loaches (formerly included in Balitoridae; Sawada, 1982); and Vaillantellidae, containing the vaillantellid loaches (formerly considered a nemacheilid; Sawada, 1982) or a botiid (Nalbant and Bǎnǎrescu, 1977; Siebert, 1987). We have matched the Saitoh et al. (2006) dataset as closely as possible with four independent nuclear genes for the same taxa (with few exceptions). The goal was to compare the phylogenetic signal of additional, independent genes with that of the mitochondrial genome, and to examine congruence in phylogenetic signal for relationships among these clades. Our working hypothesis was that analysis of both mitochondrial and nuclear data would result in isomorphic phylogenetic relationships of the major clades of cypriniform fishes. This hypothesis was not supported, as some sister group relationships varied depending upon the data examined. Potential explanations include (1) different evolutionary histories of the included nuclear genes and the mitochondrial genome, (2) rate variation among included genes, (3) limited taxon sampling, and (4) insufficient characters.

The fundamental question in cypriniform systematics is the phylogenetic relationships of the included families. At present, the eight cypriniform families are considered to form two major clades, the Cyprinoidea, containing the Cyprinidae, and the Cobitoidea, containing all remaining families (Siebert, 1987; Saitoh et al., 2006; Šlechtová et al., 2007); however, there has been little support in previous analyses for the relationships of the Cyprinidae, Catostomidae, Gyrinocheilidae (Conway et al., in press; Simons and Gidmark, in press). There is strong support for monophyly of the remaining taxa, herein referred to as the loach clade, based on morphological (Sawada, 1982; Siebert, 1987) and molecular data (Saitoh et al., 2006; Šlechtová et al., 2007). This pattern is continued in our analyses, the instability at the base

of the cypriniform tree is reflected by each of the four nuclear genes added to the Saitoh et al. (2006) dataset. The traditional division of cypriniform fishes into a monophyletic Cobitoidea and Cyprinoidea was recovered only by RAG1 (the same gene sequenced by Šlechtová et al. (2007)) and analysis of all nuclear genes together (Figs. 1 and 2; Table 3). Rhodopsin recovered a paraphyletic Cyprinoidea containing a monophyletic Cobitoidea; in contrast, GH and IRBP recovered a paraphyletic Cobitoidea containing a monophyletic Cyprinoidea. The latter situation was also observed in the combined analysis of nuclear and mitochondrial genes (Fig. 3). However, such a conflict can only be resolved in light of taxonomic congruence from independent gene/genome trees (Mickevich, 1978; Miyamoto and Fitch, 1995; Chen et al., 2003). In our study, both of the clades (Cobitoidea and Cyprinoidea) were repeatedly recovered in the majority of gene/genome data analyses (Table 3). Reciprocal monophyly of these two cypriniform superfamilies should be valid, at least from currently available data.

The family Catostomidae contains the suckers, a group largely restricted to North America. There is a substantial amount of support for the monophyly of this clade including morphological characters (summarized by Conway et al. (in press)) and molecular data (Saitoh et al., 2006; Šlechtová et al., 2007; this study). They are often considered the sister taxon to the rest of Cobitoidea; however, IRBP recovered them as sister to all Cypriniformes, a result also observed in the combined analysis of nuclear and mitochondrial genes; however, this result may not be reliable as it is never repeated in independent datasets and analyses.

The family Gyrinocheilidae contains the algae eaters, a group of three species that are widespread in Southeast Asia. The algae eaters have a very interesting morphology. They lack teeth (all other cypriniform fishes have teeth only on the fifth ceratobranchial bone) and have a peculiar sucker-like mouth that is lined with keratinous rasps (Benjamin, 1986). The phylogenetic position of the gyrinocheilids is also unstable. Combined analysis of nuclear and mitochondrial genes places *Gyrinocheilus* sister to the loach clade within the Cobitoidea; however, GH places *Gyrinocheilus* as sister to the remaining Cypriniformes. A clear picture of the phylogenetic position of Gyrinocheilidae does not emerge from our data especially from separate analyses of gene/genome. Relationships of this genus will require further evaluation with additional independent gene data and evaluation of morphological data.

The family Cyprinidae is one of the largest families of freshwater fishes and contains most of the taxa included in Cypriniformes. Cyprinidae is recovered as a monophyletic group in most of our analyses; however, analysis of rhodopsin sequences do not recover a monophyletic Cyprinidae (Table 3) which is rendered paraphyletic by the Cobitoidea. In this case, the cyprinine clade plus Clade IV (*Danio* and *Esomus*) of cyprinids together is sister to the Cobitoidea, albeit with low support.

Analysis of RAG1 sequences produces similar results to relationships derived from mitochondrial sequences (Saitoh et al., 2006) and previous studies with RAG1 sequences (Šlechtová et al., 2007), though the other nuclear loci do not always corroborate the RAG1 results. These conflicts could be explained as the result of different evolutionary histories of the included nuclear genes and the mitochondrial genome; however, examination of unrooted topologies illustrate similarities among all analyses. In all cases, regardless of gene or analysis (equal-weighting or RY coding) the leuciscine cyprinid group and the cyprinine cyprinid group are closely related as is the Catostomidae and the loach clade. What does differ between all the analyses are the position of gyrincheilids and the location of the root. Gyrincheilids are variously positioned on an internal branch between the loach/catostomid clade and the cyprinid clade, on the loach clade, or on the catostomid clade. Similarly, the outgroups may root on the leucisine cyprinids,

the gyrinocheilids, the catostomids, or another internal branch. One logical explanation for at least some of the differences observed herein between these two genomes has to do with branch lengths across the trees for alternate genes or genomes. While we have employed a model-based approach for all analyses, our necessary restriction in this comparative study to roughly 1% (~50 species of an estimated 5000 species) of the predicted number of species in the order may, infact, be responsible for the observed incongruencies observed because of long-branch attraction or taxon sampling. It should be noted that incorrect model estimations due to inadequate species sampling could result in biased tree topologies. Therefore, intense taxonomic sampling can not only correct this systematic error of estimations but also break up long branches due to heterogeneous rates of gene evolution across taxa, thereby increasing the accuracy of our phylogenetic estimation (Rannala et al., 1998; Huelsenbeck and Lander, 2003). However, several authors (Hillis et al., 1994; Graybeal, 1994; Rannala et al., 1998; Sullivan et al., 1999) have also shown that increases in the number of characters can increase resolution and support for a phylogeny. That is, by increasing character data, such as the number of genes or nucleotides, it should reduce stochastic error or character sampling bias and maximize congruence of all relevant characters to obtain the preferred hypothesis (Chen et al., 2004; Philippe et al., 2004; Delsuc et al., 2005).

Over the last two decades a plethora of parametric and nonparametric algorithms have been advocated for the inference of historical relationships of taxa. Ideally, one hoping to examine the phylogenetic relationships of a group would have access to all taxa from all clades, even extinct taxa if fossils are available, such that accurate hypotheses of relationships could be deduced from as much data as is available. Likewise, given the technological advances in molecular biology, an abundance of data should be available and most would involve simple homology assessment to derive phylogenies. As identified by Swofford et al. (1996); Heath et al. (2008); and Mayden et al. (2008), such an ideal situation rarely exists and researchers must infer relationships based on only a subsampling of available taxa and characters. The question remains then as to how effective is the subsampling and union of different data types in inferences of the tree of life? As identified in the above listed studies, several important considerations must be given to any analysis for researchers to most accurately infer relationships. Accurate phylogenetic reconstructions depend upon: (1) accurate analytical methods, (2) selection of appropriate types of and enough character data (e.g., morphological, molecular, behavioral) for reliable inference, and (3) appropriate selection of taxa; that is, the appropriate selection of taxa for the question at hand. Herein, we are working with roughly the same taxa and using the same analyses, varying only the types of characters used in phylogeny reconstruction, thereby constraining two of these three considerations.

The ease with which DNA sequence data may be obtained, combined with the improved models and model-based analyses, has led to rapid and widespread adoption of these methods by the community. However, for a variety of reasons, many researchers have nonetheless limited analyses to only a subset of taxa for a proposed clade, often because of unacceptable computation time, availability of specimens, and/or limited funds to collect large numbers of homologous sequences. This has resulted in a culture of scientists focusing on relatively few taxa with an abundance of character data. Analyses of relationships among relatively few taxa based on complete or nearly complete genome data represent an excellent example of one extreme (e.g., Inoue et al., 2001; Miya et al., 2003; Mabuchi et al., 2007).

In recent years, algorithms have improved computation time for large numbers of taxa (Huelsenbeck and Ronquist, 2001; Stamatakis, 2006). With this flexibility, one may then ask, when given a choice, which of the two possible variables should be increased, taxa or character data, to increase the accuracy of the inferred phylogeny? This question has been the focus of a number of studies and debate (Hillis et al., 2003; Rosenberg and Kumar, 2003; Rokas et al., 2003; Cummings and Meyer, 2005; Rokas and Carroll, 2005; Hedtke et al., 2006; Gatesy et al., 2007; Heath et al., 2008; Mayden et al., 2008). Evidence usually supports increasing taxon sampling, even at the expense of great quantities of character data, for improved accuracy of topologies. In simulation studies, increased taxon sampling appears to be more consequential than increasing the number of characters (Hillis, 1996). Several authors have also agreed that the addition of species in analyses results in more accurate estimates of relationships (Lecointre et al., 1993; Hillis, 1996, 1998; Graybeal, 1998; Rannala et al., 1998; Zwickl and Hillis, 2002; Pollock et al., 2002; Poe, 2003; DeBry, 2005; Hedtke et al., 2006). Furthermore, studies have attributed problematic reconstructions and poorly resolved trees to a limited number of taxa (Bremer et al., 1999; Johnson, 2001; Lin et al., 2002; Braun and Kimball, 2002; Chen et al., 2003; Sorenson et al., 2003; Albrecht et al., 2007).

The essential problems with focusing only on increasing characters at the expense of taxa involves complications with estimates of unobserved changes or transformations in a tree— consequently poor estimates of evolutionary models or a resulting matrix that precludes parsimony from arriving at a correct solution. With too few taxa it is difficult or impossible to accurately estimate parameters for evolutionary models as there will be too many unobserved changes in a matrix (Felsenstein, 1978; Hendy and Penny, 1989; Debry, 1992; Huelsenbeck and Hillis, 1993; Yang, 1994; Huelsenbeck, 1995; Gascuel et al., 2001; Huelsenbeck and Lander, 2003; Susko et al., 2004). Serious complications include either long-branch attraction (Felsenstein, 1978) or a nodal-density effect (Gojobori et al., 1982; Fitch and Bruschi, 1987; Fitch and Beintema, 1990; Bruno and Halpern, 1999; Hugall and Lee, 2007), or both. Here, a limited sampling of species results in an artificial accumulation of apomorphies possessed by species (ancestral or descendant) because taxa are missing from intervening nodes that would presumably "break up" branches and more realistically disperse character change in the phylogeny (apomorphies and homoplasy) (Wiens, 2005). Long-branch attraction results from an accumulation of phylogenetic noise or homoplasy in two or more non-adjacent taxa that is interpreted as an accumulation of homologous characters between two closely related species. In either case, a restriction of taxon sampling, even with a limited number of characters, can result in phylogenetic noise (homoplasy via convergences, reversals, or substitutions) overwhelming the phylogenetic signal. This study was an attempt to test the mitogenome approach by comparing that tree to a tree determined from an unrelated set of nuclear loci. While the results obtained herein were in general agreement regarding relationships among Cypriniformes, questions regarding relationships remain. Obviously, the most appropriate method to address these unresolved issues is a more comprehensive sampling of the diversity present within this important group of fishes. Future efforts, while likely to include the same sets of genetic markers, should include a much denser representation of the taxonomic diversity of cypriniform fishes.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.12.015.

References

Albrecht, C., Kuhn, K., Streit, B., 2007. A molecular phylogeny of Planorboidea (Gastropoda: Pulmonata): insights from enhanced taxon sampling. Zool. Scripta. 36, 27–39.

Benjamin, M., 1986. The oral sucker of Gyrinocheilus aymonieri (Teleostei: Cypriniformes). J. Zool., Lond. (B) 1, 211–254.

Bernardi, G., D'Onofrio, G., Caccio, S., Bernardi, G., 1993. Molecular phylogeny of bony fishes, based on the amino acid sequence of the growth hormone. J. Mol. Evol. 37, 644–649.

Bielawski, J.P., Gold, J.R., 2001. Phylogenetic relationships of cyprinid fishes in subgenus Notropis inferred from nucleotide sequences of the mitochondrially encoded cytochrome b gene. Copeia 2001, 656–667.

Braun, E.L., Kimball, R.T., 2002. Examining basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. Syst. Biol. 51, 614–625.

Bremer, B., Jansen, R.K., Oxelman, B., Backlund, M., Lantz, H., Kim, K.J., 1999. More characters or more taxa for a robust phylogeny—case study from the coffee family (Rubiaceae). Syst. Biol. 48, 413–435.

Briolay, J., Galtier, N., Brito, R.M., Bouvet, Y., 1998. Molecular phylogeny of Cyprinidae inferred from cytochrome b DNA sequences. Mol. Phylogenet. Evol. 9, 100–108.

Broughton, R.E., Gold, J.R., 2000. Phylogenetic relationships in the North American cyprinid genus Cyprinella (Actinopterygii: Cyprinidae) based on sequences of the mitochondrial ND2 and ND4L genes. Copeia 2000, 1–10.

Bruno, W.J., Halpern, A.L., 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Mol. Biol. Evol. 16, 564–566.

Chen, W.-J., Bonillo, C., Lecointre, G., 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Mol. Phylogenet. Evol. 26, 262–288.

Chen, W.-J., Ortí, G., Meyer, A., 2004. Novel evolutionary relationship among four fish model systems. Trends Genet. 20, 424–431.

Chen, W.-J., Ruiz-Carus, R., Ortí, G., 2007. Relationships among four genera of mojarras (Teleostei: Perciformes: Gerreidae) from the western Atlantic and their tentative placement among percomorph fishes. J. Fish Biol. 70 (sb), 202–218.

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, 25–134.

Clements, M.D., Bart Jr., H.L., Hurley, D.L., 2004. Isolation and characterization of two distinct growth hormone cDNAs from the tetraploid smallmouth buffalofish (Ictiobus bubalus). Gen. Comparat. Endocrinol. 136, 411–418.

Conway, K.W., Hirt, M.V., Yang, L., Mayden, R.L., Simons, A.M., in press. Cypriniformes: systematics and paleontology. In: Schultz, H.-P., Wilson, M.V.H., Nelson, J.S. (Eds.), Origin and Phylogenetic Interrelationships of Teleosts, Verlag Dr. Freidrich Feil, Munich.

Cunha, C., Mesquita, N., Dowling, T.E., Gilles, A., Coelho, M.M., 2002. Phylogenetic relationships of Eurasian and American cyprinids using cytochrome b sequences. J. Fish Biol. 61, 929–944.

Cummings, M.P., Meyer, A., 2005. Magic bullets and golden rules: data sampling in molecular phylogenetics. Zoology 108, 329–336.

David, L., Blum, S., Feldman, M.W., Lavi, U., Hillel, J., 2003. Recent duplication of the common carp (Cyprinus carpio L.) genome as revealed by analyses of microsatellite loci. Mol. Biol. Evol. 20, 1425–1434.

DeBry, R.W., 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. Mol. Biol. Evol. 9, 537–551.

DeBry, R.W., 2005. The systematic component of phylogenetic error as a function of taxonomic sampling under parsimony. Syst. Biol. 54, 432–440.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.

Dettaï, A., Lecointre, G., 2004. In search of notothenioid (Teleostei) relatives. Antarct. Sci. 16, 71–85.

Dettaï, A., Lecointre, G., 2005. Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. C.R. Bios. 328, 674–689.

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). Mol. Phylogenet. Evol. 48, 258–269.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Dowling, T.E., Naylor, G.J.P., 1997. Evolutionary relationships of minnows in the genus Luxilus (Teleostei: Cyprinidae) as determined from cytochrome b sequences. Copeia 1997, 758–765.

Durand, J.-D., Tsigenopoulos, C.S., Ünlü, E., Berrebi, P., 2002. Phylogeny and biogeography of the family Cyprinidae in the Middle East inferred from cytochrome b DNA—evolutionary significance of this region. Mol. Phylogenet. Evol. 22, 91–100.

Felsenstein, J., 1978. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. 16, 183–196.

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

Fitch, W.M., Beintema, J.J., 1990. Correcting parsimonious trees for unseen nucleotide substitutions: the effect of dense branching as exemplified by ribonuclease. Mol. Biol. Evol. 7, 438–443.

Fitch, W.M., Bruschi, M., 1987. The evolution of prokaryotic ferredoxins—with a general method correcting for unobserved substitutions in less branched lineages. Mol. Biol. Evol. 4, 381–394.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299.

Gascuel, O., Bryant, D., Denis, F., 2001. Strengths and limitations of the minimum evolution principle. Syst. Biol. 50, 621–627.

Gatesy, J., DeSalle, R., Wahlberg, N., 2007. How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. Syst. Biol. 56, 355–363.

Gojobori, T., Ishii, K., Nei, M., 1982. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. J. Mol. Evol. 18, 414–423.

Goldman, N., 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36, 182–198.

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

Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47, 9–17.

He, S.P., Liu, H.Z., Chen, Y.Y., Kuwahara, M., Nakajima, T., Zhong, Y., 2004. Molecular phylogenetic relationships of Eastern Asian Cyprinidae (Pisces: Cypriniformes) inferred from cytochrome b sequences. Sci. China Ser. C Life Sci. 47, 130–138.

He, S.P., Mayden, R.L., Wang, X.Z., Wang, W., Tang, K.L., Chen, W.-J., Chen, Y.Y., 2008a. Molecular phylogenetics of the family Cyprinidae (Actinopterygii: Cypriniformes) as evidenced by sequence variation in the first intron of S7 ribosomal protein-coding gene: further evidence from a nuclear gene of the systematic chaos in the family. Mol. Phylogenet. Evol. 46, 818–829.

He, S.P., Gu, X., Mayden, R.L., Chen, W.-J., Conway, K.W., Chen, Y.Y., 2008b. Phylogenetic position of the enigmatic genus Psilorhynchus (Ostariophysi: Cypriniformes): evidence from the mitochondrial genome. Mol. Phylogenet. Evol. 47, 419–425.

Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46, 239–257.

Hedtke, S.M., Townsend, T.M., Hillis, D.M., 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55, 522–529.

Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Hillis, D.M., 1996. Inferring complex phylogenies. Nature 383, 130–131.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47, 3–8.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.

Hillis, D.M., Huelsenbeck, J.P., Cunningham, C.W., 1994. Application and accuracy of molecular phylogenies. Science 264, 671–677.

Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J., 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52, 124–126.

Holcroft, N.I., 2005. A molecular analysis of the interrelationships of tetraodontiform fishes (Acanthomorpha: Tetraodontiformes). Mol. Phylogenet. Evol. 34, 525–544.

Huelsenbeck, J.P., 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44, 17–48.

Huelsenbeck, J.P., Hillis, D.M., 1993. Success of phylogenetic methods in the 4-taxon case. Syst. Biol. 42, 247–264.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Huelsenbeck, J.P., Lander, K.M., 2003. Frequent inconsistency of parsimony under a simple model of cladogenesis. Syst. Biol. 52, 641–648.

Hugall, A.F., Lee, M.S.Y., 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61, 2293–2307.

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2001. A mitogenomic perspective on the basal teleostean phylogeny: higher-level relationships with longer DNA sequences. Mol. Phylogenet. Evol. 20, 275–285.

Johnson, K.P., 2001. Taxon sampling and the phylogenetic position of passeriformes: evidence from 916 avian cytochrome b sequences. Syst. Biol. 50, 128–136.

Kawakita, A., Sota, T., Ascher, J.S., Ito, M., Tanaka, H., Kato, M., 2003. Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (Bombus). Mol. Biol. Evol. 20, 87–92.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86, 6196–6200.

Larhammar, D., Risinger, C., 1994. Molecular genetic aspects of tetraploidy in the common carp Cyprinus carpio. Mol. Phylogenet. Evol. 3, 59–68.

Lecointre, G., Philippe, H., Vân Lê, H.L., Le Guyader, H., 1993. Species sampling has a major impact on phylogenetic inference. Mol. Phylogenet. Evol. 2, 205–224.

Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11, 605–612.

Lin, Y.H., McLenachan, P.A., Gore, A.R., Phillips, M.J., Ota, R., Hendy, M.D., Penny, D., 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol. Biol. Evol. 19, 2060–2070.

López, J.A., Chen, W.-J., Ortí, G., 2004. Esociform phylogeny. Copeia 2004, 449–464.

Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. 49, 628–651.

Mabuchi, K., Miya, M., Azuma, Y., Nishida, M., 2007. Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. BMC Evol. Biol. 7, 10.

Machordom, A., Doadrio, I., 2001. Evidence of a Cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (Luciobarbus, Cyprinidae). Mol. Phylogenet. Evol. 18, 252–263.

Mayden, R.L., 1992. The fundamentals of phylogenetic systematics. In: Mayden, R.L. (Ed.), Systematics, Historical Ecology, and North American Freshwater Fishes. Stanford University Press, Palo Alto, pp. 114–185.

Mayden, R.L., Simons, A.M., Wood, R.M., Harris, P.M., Kuhajda, B.R., 2006. Molecular systematics and classification of North American notropin shiners and minnows (Cypriniformes: Cyprinidae). In: Lozano-Vilano, M.L., Contreras-Balderas, A.J. (Eds.), Studies of North American Desert Fishes in Honor of E. P. (Phil) Pister. Conservationist. Universidad Autónoma de Nuevo León, Monterrey, pp. 72–101.

Mayden, R.L., Tang, K.L., Conway, K.W., Freyhof, J., Chamberlain, S., Haskins, M., Schneider, L., Sudkamp, M., Wood, R.M., Agnew, M., Bufalino, A., Sulaiman, Z., Miya, M., Saitoh, K., He, S.P., 2007. Phylogenetic relationships of Danio within the order Cypriniformes: a framework for comparative and evolutionary studies of a model species. J. Exp. Zool. B Mol. Dev. Evol. 308B, 642–654.

Mayden, R.L., Tang, K.L., Wood, R.M., Chen, W.-J., Agnew, M.K., Conway, K.W., Yang, L., Simons, A.M., Bart, H.L., Harris, P.M., Li, J.B., Wang, X.Z., Saitoh, K., He, S.P., Liu, H.Z., Chen, Y.Y., Nishida, M., Miya, 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. J. Syst. Evol. 46, 424–438.

Mickevich, M.F., 1978. Taxonomic congruence. Syst. Zool. 27, 143–158.

Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M., Nishida, M., 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phylogenet. Evol. 26, 121–138.

Miya, M., Saitoh, K., Wood, R.M., Nishida, M., Mayden, R.L., 2006. New primers for amplifying and sequencing the mitochondrial ND4/ND5 gene region of the Cypriniformes (Actinopterygii: Ostariophysi). Ichthyol. Res. 53, 75–81.

Miyamoto, M.M., Fitch, W.M., 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44, 64–76.

Mullis, K.B., Faloona, F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155, 335–350.

Moore, W.S., 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49, 718–726.

Nalbant, T.T., Bănărescu, P.M., 1977. Vaillantellinae, a new subfamily of Cobitidae (Pisces, Cypriniformes). Zool. Meded. 52, 99–105.

Nylander, J.A.A., 2004. MrModeltest v2.2. Evolutionary Biology Centre, Uppsala University, Uppsala.

Oakley, T.H., Phillips, R.B., 1999. Phylogeny of salmonine fishes based on growth hormone introns: Atlantic (Salmo) and Pacific (Oncorhynchus) salmon are not sister taxa. Mol. Phylogenet. Evol. 11, 381–393.

Ohno, S., Muramoto, J., Christian, L., Atkin, N.B., 1967. Diploid-tetraploid relationship among Old World members of the fish family Cyprinidae. s s s 23, 1–9.

Palumbi, S.R., 1996. Nucleic acids II: the polymerase chain reaction. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, pp. 205–247.

Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W.H., Casane, D., 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol. Biol. Evol. 21, 1740–1752.

Poe, S., 2003. Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. Syst. Biol. 52, 423–428.

Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51, 664–671.

Rajesh, R., Majumdar, K.C., 2007. A comparative account of the structure of growth hormone encoding gene and genetic interrelationships is six species of genus Labeo. Fish. Physiol. Biochem. 33, 311–333.

Raley, M.E., Wood, R.M., 2001. Molecular systematics of members of the Notropis dorsalis species group (Actinopterygii: Cyprinidae). Copeia 2001, 638–645.

Rannala, B., Huelsenbeck, J.P., Yang, Z.H., Nielsen, R., 1998. Taxon sampling and the accuracy of large phylogenies. Syst. Biol. 47, 702–710.

Rokas, A., Carroll, S.B., 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. 22, 1337–1344.

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.

Rosenberg, M.S., Kumar, S., 2003. Taxon sampling, bioinformatics, and phylogenomics. Syst. Biol. 52, 119–124.

Rüber, L., Britz, R., Kullander, S.O., Zardoya, R., 2004. Evolutionary and biogeographic patterns of the Badidae (Teleostei: Perciformes) inferred from mitochondrial and nuclear DNA sequence data. Mol. Phylogenet. Evol. 32, 1010–1022.

Rüber, L., Kottelat, M., Tan, H.H., Ng, P.K.L., Britz, R., 2007. Evolution of miniaturization and the phylogenetic position of Paedocypris, comprising the world's smallest vertebrate. BMC Evol. Biol. 7, 38.

Rambaut, A., 1996. Se-Al: Sequence alignment editor version 1.0 α1. Department of Zoology, University of Oxford, Oxford. Available from: <http://tree.bio.ed.ac.uk/software/seal/>.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A., 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.

Saitoh, K., Miya, M., Inoue, J.G., Ishiguro, N.B., Nishida, M., 2003. Mitochondrial genomics of ostariophysan fishes: perspectives on phylogeny and biogeography. J. Mol. Evol. 56, 464–472.

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 towards resolution of higher-level relationships of the world's largest freshwater-fish clade based on 59 whole mitogenome sequences. J. Mol. Evol. 63, 826–841.

Sawada, Y., 1982. Phylogeny and zoogeography of the superfamily Cobitoidea (Cyprinoidei, Cypriniformes). Mem. Fac. Fish. Hokkaido Univ. 28, 65–223.

Schmidt, T.R., Dowling, T.E., Gold, J.R., 1994. Molecular systematics of the genus Pimephales (Teleostei: Cyprinidae). Southwest. Nat. 39, 241–248.

Schmidt, T.R., Bielawski, J.P., Gold, J.R., 1998. Molecular phylogenetics and evolution of the cytochrome b gene in the cyprinid genus Lythrurus (Actinopterygii: Cypriniformes). Copeia 1998, 14–22.

Schönhuth, S., De Sostoa, A., Martínez, E., Doadrio, I., 2001. Southern Mexican minnows of the genus Notropis (Actinopterygii, Cyprinidae): genetic variation, phylogenetic relationships and biogeographical implications. Biochem. Syst. Ecol. 29, 359–377.

Schönhuth, S., Doadrio, I., 2003. Phylogenetic relationships of Mexican minnows of the genus Notropis (Actinopterygii: Cyprinidae). Biol. J. Linn. Soc. 80, 323–337.

Schönhuth, S., Doadrio, I., Mayden, R.L., 2007. A biogeographic perspective on the phylogeny of Mexican cyprinids (Actinopterygii, Cyprinidae). In: Lozano-Vilano, M.L., Contreras-Balderas, A.J. (Eds.), Studies of North American Desert Fishes in Honor of E. P. (Phil) Pister. Conservationist. Universidad Autónoma de Nuevo León, Monterrey, pp. 102–124.

Schönhuth, S., Doadrio, I., Dominguez-Dominguez, O., Hillis, D.M., Mayden, R.L., 2008. Molecular evolution of southern North American Cyprinidae (Actinopterygii), with the description of the new genus Tampichthys from central Mexico. Mol. Phylogenet. Evol. 47, 729–756.

Siebert, D.J., 1987. Interrelationships among families of the order Cypriniformes (Teleostei)—Unpubl. Ph.D. Thesis, City University of New York, New York, NY.

Simons, A.M., 2004. Phylogenetic relationships in the genus Erimystax (Actinopterygii: Cyprinidae) based on the cytochrome b gene. Copeia 2004, 351–356.

Simons, A.M., Gidmark, N.J., in press. Systematics and phylogenetic relationships of the Cypriniformes. In: Grande, T.C. (Ed.), A Comprehensive Review of Gonorynchiformes and of Ostariophysan Relationships. Scientific Publishers Inc., Enfield.

Simons, A.M., Mayden, R.L., 1997. Phylogenetic relationships of the creek chubs and the spine-fins: an enigmatic group of North American cyprinid fishes (Actinopterygii: Cyprinidae). Cladistics 13, 187–205.

Simons, A.M., Mayden, R.L., 1998. Phylogenetic relationships of the western North American phoxinins (Actinopterygii: Cyprinidae) as inferred from mitochondrial 12S and 16S RNA sequences. Mol. Phylogenet. Evol. 9, 308–329.

Simons, A.M., Mayden, R.L., 1999. Phylogenetic relationships of North American cyprinids and assessment of homology of the open posterior myodome. Copeia 1999, 13–21.

Simons, A.M., Knott, K.E., Mayden, R.L., 2000. Assessment of monophyly and evolution of nest association in the minnow genus Pteronotropis (Teleostei: Cyprinidae). Copeia 2000, 1068–1075.

Simons, A.M., Berendzen, P.B., Mayden, R.L., 2003. Molecular systematics of North American phoxinin genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. J. Linn. Soc. 139, 63–80.

Slade, R.W., Moritz, C., Heideman, A., 1994. Multiple nuclear-gene phylogenies: application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. Mol. Biol. Evol. 11, 341–356.

Šlechtová, V., Bohlen, J., Tan, H.H., 2007. Families of Cobitoidea (Teleostei; Cypriniformes) as revealed from nuclear genetic data and the position of the mysterious genera Barbucca, Psilorhynchus, Serpenticobitis and Vaillantella. Mol. Phylogenet. Evol. 44, 1358–1365.

Šlechtová, V., Bohlen, J., Perdices, A., 2008. Molecular phylogeny of the freshwater fish family Cobitidae (Cypriniformes: Teleostei): delimitation of genera, mitochondrial introgression and evolution of sexual dimorphism. Mol. Phylogenet. Evol. 47, 812–831.

Sorenson, M.D., Oneal, E., García-Moreno, J., Mindell, D.P., 2003. More taxa, more characters: the hoatzin problem is still unresolved. Mol. Biol. Evol. 20, 1484–1498.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web-servers. Syst. Biol. 57, 758–771.

Sullivan, J., Swofford, D.L., Naylor, G.J.P., 1999. The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. Mol. Biol. Evol. 16, 1347–1356.

Sullivan, J.P., Lundberg, J.G., Hardman, M., 2006. A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. Mol. Phylogenet. Evol. 41, 636–662.

Susko, E., Inagaki, Y., Roger, A.J., 2004. On inconsistency of the neighbor-joining, least squares, and minimum evolution estimation when substitution processes are incorrectly modeled. Mol. Biol. Evol. 21, 1629–1642.

Swofford, D.L., 2002. PAUP∗. Phylogenetic Analysis Using Parsimony (∗ and Other Methods). Version 4.0b10, fourth ed. Sinauer Associates, Sunderland.

Swofford, D.L., Olsen, G.P., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, pp. 407–514.

Tang, Q.Y., Liu, H.Z., Mayden, R.L., Xiong, B.X., 2006. Comparison of evolutionary rates in the mitochondrial DNA cytochrome *b* gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes). Mol. Phylogenet. Evol. 39, 347–357.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Uyeno, T., Smith, G.R., 1972. Tetraploid origin of the karyotype of catostomid fishes. Science 175, 644–646.

Vawter, L., Brown, W.M., 1986. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science 234, 194–196.

Wang, X.Z., Wang, J., He, S.P., Mayden, R.L., 2007. The complete sequence of the Chinese hook snout carp (*Opsariichthys bidens*) mitochondrial genome and an alternative pattern of mitogenomic evolution in vertebrate. Gene 399, 11–19.

Wiens, J., 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Syst. Biol. 54, 731–742.

Wiley, E.O., 1981. Phylogenetics: The Theory and Practice of Phylogenetic Systematics. Wiley-Interscience, New York.

Yang, Z.H., 1994. Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. Syst. Biol. 43, 329–342.

Yu, X., Zhou, T., Li, K., Li, Y., Zhou, M., 1987. On the karyosystematics of cyprinid fishes and a summary of fish chromosome studies in China. Genetica 72, 225–236.

Zaragüeta-Bagils, R., Lavoué, S., Tillier, A., Bonillo, C., Lecointre, G., 2002. Assessment of otocephalan and protacanthopterygian concepts in the light of multiple molecular phylogenies. C.R. Biol. 325, 1191–1207.

Zardoya, R., Doadrio, I., 1999. Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. J. Mol. Evol. 49, 227–237.

Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588–598.