Journal of Systematics and Evolution 46 (3): 424–438 (2008) doi: 10.3724/SP.J.1002.2008.08062
(formerly Acta Phytotaxonomica Sinica) http://www.plantsystematics.com

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

[1]Richard L. MAYDEN* [1]Kevin L. TANG [1]Robert M. WOOD [1]Wei-Jen CHEN [1]Mary K. AGNEW
[1]Kevin W. CONWAY [1]Lei YANG [2]Andrew M. SIMONS [3]Henry L. BART [4]Phillip M. HARRIS
[5]Junbing LI [5]Xuzhen WANG [6]Kenji SAITOH [5]Shunping HE [5]Huanzhang LIU
[5]Yiyu CHEN [7]Mutsumi NISHIDA [8]Masaki MIYA

[1](*Department of Biology, Saint Louis University,* St. Louis, MO 63103, USA)
[2](*Department of Fisheries, Wildlife, and Conservation Biology, Bell Museum of Natural History, University of Minnesota,* St. Paul, MN 55108, USA)
[3](*Department of Ecology and Evolutionary Biology, Tulane University,* New Orleans, LA 70118, USA)
[4](*Department of Biological Sciences, The University of Alabama,* Tuscaloosa, AL 35487, USA)
[5](*Laboratory of Fish Phylogenetics and Biogeography, Institute of Hydrobiology, Chinese Academy of Sciences,* Wuhan 430072, China)
[6](*Tohoku National Fisheries Research Institute, Fisheries Research Agency,* Miyagi 985-0001, Japan)
[7](*Ocean Research Institute, University of Tokyo,* Tokyo 164-8639, Japan)
[8](*Department of Zoology, Natural History Museum & Institute,* Chiba 260-8682, Japan)

Abstract The phylogenetic relationships of species are fundamental to any biological investigation, including all evolutionary studies. Accurate inferences of sister group relationships provide the researcher with an historical framework within which the attributes or geographic origin of species (or supraspecific groups) evolved. Taken out of this phylogenetic context, interpretations of evolutionary processes or origins, geographic distributions, or speciation rates and mechanisms, are subject to nothing less than a biological experiment without controls. Cypriniformes is the most diverse clade of freshwater fishes with estimates of diversity of nearly 3,500 species. These fishes display an amazing array of morphological, ecological, behavioral, and geographic diversity and offer a tremendous opportunity to enhance our understanding of the biotic and abiotic factors associated with diversification and adaptation to environments. Given the nearly global distribution of these fishes, they serve as an important model group for a plethora of biological investigations, including indicator species for future climatic changes. The occurrence of the zebrafish, *Danio rerio*, in this order makes this clade a critical component in understanding and predicting the relationship between mutagenesis and phenotypic expressions in vertebrates, including humans. With the tremendous diversity in Cypriniformes, our understanding of their phylogenetic relationships has not proceeded at an acceptable rate, despite a plethora of morphological and more recent molecular studies. Most studies are pre-Hennigian in origin or include relatively small numbers of taxa. Given that analyses of small numbers of taxa for molecular characters can be compromised by peculiarities of long-branch attraction and nodal-density effect, it is critical that significant progress in our understanding of the relationships of these important fishes occurs with increasing sampling of species to mitigate these potential problems. The recent Cypriniformes Tree of Life initiative is an effort to achieve this goal with morphological and molecular (mitochondrial and nuclear) data. In this early synthesis of our understanding of the phylogenetic relationships of these fishes, all types of data have contributed historically to improving our understanding, but not all analyses are complementary in taxon sampling, thus precluding direct understanding of the impact of taxon sampling on achieving accurate phylogenetic inferences. However, recent molecular studies do provide some insight and in some instances taxon sampling can be implicated as a variable that can influence sister group relationships. Other instances may also exist but without inclusion of more taxa for both mitochondrial and nuclear genes, one cannot distinguish between inferences being dictated by taxon sampling or the origins of the molecular data.

Key words Cobitoidea, Cypriniformes, Cyprinoidea, taxon sampling, Tree of Life.

Received: 27 April 2008 Accepted: 9 May 2008
* Author for correspondence. E-mail: cypriniformes@gmail.com; Tel.: 1-314-977-3910; Fax: 1-314-977-3658.

Nothing in Biology Makes Sense Except in the Light of Evolution.
Theodosius Dobzhansky. 1973.
The American Biology Teacher

Phylogenetic reconstruction is fundamental to comparative biology research … as the phylogeneticists' conclusions (i.e., their phylogenetic inferences) become the comparative biologists' assumptions. Consequently, the generation of robust phylogenetic hypotheses and the understanding of the factors influencing accuracy in phylogenetic reconstruction are crucial to evolutionary hypothesis testing.
Antonis Rokas and Sean B. Carroll (2005: 1337).

This famous quote from Dobzhansky emerged over a century after Darwin's transformational treatise on descent with modification changed the face of comparative biology. During the long hiatus following Darwin's hypothesis, much controversy emerged about the idea of evolution and speciation. This persisted until advances in genetics and population biology made these disciplines which scientists could use to provide first order, hypothesis-driven explanations for evolution and speciation. While dominating the field of evolutionary biology for decades, neither this discipline nor the related disciplines of taxonomy or systematics could offer a satisfactory theoretical framework for reconstructing historical patterns of speciation or evolutionary mechanisms underlying the tree of life. This held true even after Hennig's methodology, now a fundamental part of phylogenetic systematics, had already been published in "*Grundzüge einer Theorie der phylogenetischen Systematik"* (1950) (the translated English version "*Phylogenetic Systematics*" that received a much wider international distribution was not published until 1966). Integral in Hennig's work was the evaluation of both population and species level evolution of traits or attributes, speciation, and inheritance of these traits throughout the tree of life. His theories were quite contrary to the traditional view of systematic biology of the time and, as a discipline dealing with systematics, were not viewed as worthy of investigation by geneticists and population biologists, then forging new ideas on evolution. As with any paradigm shift (Kuhn, 1962), significant controversy and established inertia delayed an unbiased assessment and the eventual adoption of this now widely accepted philosophical and methodological transformation, offering for the first time a mechanism for researchers to reconstruct testable histories (species trees) of life. Thus, at the time of Dobzhansky's famous assessment of biology, few really knew of or understood the significance of Hennig's work. This was much like the theory of continental drift and the works of Alfred Wegener (Wegener, 1915) which had almost no impact on its field when originally published. Recognition for both of these important interdisciplinary scientists and their work did not come until after their deaths.

In the wake of the transformation of the biological community following the adoption of Hennig's theory and methods for reconstructing testable hypotheses of evolutionary relationships (i.e., phylogenies), it is clear that Dobzhansky's assessment of the essential foundations of biology requires reconsideration. Rather, as indicated in the second quote by Rokas and Carroll (2005), it is more appropriate to keep in mind that nothing in biology makes sense except in the light of the phylogenetic relationships of species, including evolution. In fact, the conclusions of any phylogenetic study of a group of organisms serve as the beginning of any other biological investigation of these same organisms. Only when a researcher is equipped with a hypothesis of the sister group relationships of targeted populations, species, or supraspecific natural (monophyletic) taxa can one address the variety of possible biological questions about any of the organisms or species, including evolutionary studies.

Given the obvious significance of hypotheses of the genealogical relationships of species, it is critical that these hypotheses reflect as closely as possible the genealogical history of the species, including sister group relationships, ancestral state reconstructions, and branch lengths derived from meaningful optimizations. Sister group relationships are critical for appropriate comparisons, biogeographic investigations, and an eventual resolution of modes of speciation. Ancestral character reconstructions and branch lengths are essential to researchers investigating many areas of evolution concerning ages of clades, rates of anagenesis and cladogenesis, tracing the descent of attributes, as well as other domains of comparative biology.

Various methods have been advocated for the inference of historical relationships of species that involve both parametric and non-parametric algorithms. In a perfect world, such a hypothesis of relationships would include all of the species within the

group and an abundance of data with straightforward homology assessment, the resulting phylogeny can be used by anyone to investigate all aspects of the species and interpret its traits and its geographic distribution within an historical framework. This perfect situation rarely exists and several important considerations must be given to any analysis, to allow a researcher to most accurately infer relationships. At least three issues must be considered for accurate phylogenetic reconstructions; these are 1) accurate analytical methods, 2) selection of appropriate and enough character data (e.g., morphology, molecular, behavioral) for reliable inference, and 3) appropriate selection of taxa for the question at hand (Swofford et al., 1996).

The computational demands inherent in phylogenetic reconstructions, representing a type of NP-complete problem (Graham & Foulds, 1982), are astounding as the number of possible resolutions increases exponentially as the number of taxa involved also increases. Historically, this prevented scientists from examining many species and a focus was placed on an increase in the number of characters for the analysis. Few morphological studies, however, come close to or exceed 100 or more characters (see Scotland et al., 2003). With the efficiency of generating molecular sequence data increasing and the cost decreasing, the ability to produce hundreds and thousands of potential characters quickly became the standard approach to assemble systematic data sets and soon the number of molecular phylogenies relative to those based on morphological data dramatically increased. With this increase in the amount of data from sequences and the character states being limited to "A-T-G-C," model-based analyses and a number of sophisticated evolutionary models have changed the face of phylogenetic systematics (Yang, 1996; Sullivan & Swofford, 1997; Stamatakis, 2006a).

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 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 data sets with large numbers of taxa (e.g., Huelsenbeck & Ronquist, 2001; Stamatakis, 2006b). 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 considerable debate (Hillis et al., 2003; Rokas et al., 2003; Rosenberg & Kumar, 2003; Cummings & Meyer, 2005; Rokas et al., 2005; Hedtke et al., 2006) and a review of this controversy is provided in Heath et al. (2008). The overwhelming evidence 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 for reaching the "true" relationships in a group (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 & Hillis, 2002; Pollock et al., 2002; Poe, 2003; DeBry, 2005; Hedtke et al., 2006). Furthermore, empirical studies have attributed problematic reconstructions and poorly resolved trees to researchers limiting analyses to an inadequate number of taxa (Bremer et al., 1999; Johnson, 2001; Lin et al., 2002; Braun & 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. If there are not enough taxa in an analysis then it is difficult to accurately estimate parameters for evolutionary models as there will be too many unobserved changes inherent in a matrix (Felsenstein, 1978; Hendy & Penny, 1989; DeBry, 1992; Huelsenbeck & Hillis, 1993; Yang, 1994; Huelsenbeck, 1995; Gascuel et al., 2001; Huelsenbeck & 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 & Bruschi, 1987; Fitch & Beintema, 1990; Bruno & Halpern, 1999; Hugall & 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.

Cypriniformes is known as the most diverse group of freshwater fishes with estimates of diversity reaching close to 3,500 species (Nelson, 2006). The family occurs in habitats ranging from lakes and rivers to small springs and streams in Eurasia, North America, and Africa. Species of this order (particularly those of the Cyprinidae) are usually perceived as having very similar morphologies (Howes, 1991), an attribute that has likely contributed to the paucity of researchers investigating their phylogenetic relationships because of suspected conserved or constrained evolution limiting the number of phylogenetically useful morphological characters. These fishes also include many commercially important species (e.g., aquarium trade, fisheries), and as model organisms in many areas of research ranging from community ecology to developmental biology. The zebrafish or zebra danio (*Danio rerio)*, goldfish (*Carassius auratus*), algae eater (*Gyrinocheilus aymonieri*), and many carp species (*Cyprinus, Hypophthalmichthys, Ctenopharyngodon*) likely represent the most familiar members of this diverse clade.

As species of Cypriniformes are diverse components in most of the freshwater habitats around the globe and serve as important model organisms in comparative research, the phylogenetic relationships of these fishes serve as a critical, historical framework aiding directed research. Unfortunately, either considerable uncertainty exists over the relationships of these fishes, even at the higher levels, or there are no data providing insight into the relationships of these fishes. Furthermore, some of the species and clades previously examined for phylogenetic relationships have been problematic in their resolution, despite efforts to increase character sampling to resolve their sister group relationships. Most notable among these are the subfamilial relationships within the families Cyprinidae, Catostomidae, Cobitidae, and Balitoridae, and the phylogenetic position of the common aquarium species, the algae eater, and relatives in the genus *Gyrinocheilus* (Gyrinocheilidae).

Herein, we examine the relationships of major clades within Cypriniformes as they relate to the impact of taxon sampling, while intentionally holding character sampling constant. We focus particularly on the historically problematic taxa identified above and how increasing taxa from 49 to 110 species in the analysis alters their phylogenetic placement and support for their sister group relationships. These comparisons are made relative to a previous analysis of these same species based on a more limited sampling of taxa (53 species) but with whole mitochondrial genomes (Saitoh et al., 2006) containing significantly more character data (14,563 bp) than used in this analysis (1497 bp). This analysis is based on sequences of exon 3 of recombination activating gene 1 (RAG1), a commonly used gene in phylogenetic relationships of gnathostome vertebrates (e.g., Groth & Barrowclough, 1999; Waddell & Shelley, 2003; San Mauro et al., 2004; Steppan et al., 2004; Krenz et al., 2005), including ray-finned fishes (e.g., López et al., 2004; Rüber et al., 2004; Holcroft, 2005; Sullivan et al., 2006; Chen et al., 2007; Mayden et al., 2007). This research is part of an ongoing international Tree of Life initiative on Cypriniformes and is aimed at furthering our understanding of not only the phylogenetic relationships of these species but also improving our understanding of effective and accurate methods for phylogeny reconstruction.

1 Methods

Taxon sampling attempted to match that of Saitoh et al. (2006). In cases where the same species was not available, a congeneric representative was chosen as a substitute, if possible. We conducted analyses on two different sets of taxa. For the first analysis, we compiled a data matrix of RAG1 sequence data to match the species from Saitoh et al. (2006; fig. 2 & table 2); this small data set included 7 outgroups and 49 cypriniform fishes. To explore the effects of taxon sampling, our second analysis added an additional 61 cypriniform taxa; these included an additional two species of catostomids, two cobitids, three botiids, four balitorids, and 50 cyprinids (one acheilognathin, one cultrin, one squaliobarbin, three gobionins, five cyprinins, 17 rasborins, and 22 leuciscins). Of those, 23 RAG1 were sequences downloaded from GenBank. Please see Table 1 for a complete list of taxa examined for this study.

RAG1 was chosen for this study as it is not part of the mitochondrial genome, thus not overlapping with the data presented in Saitoh et al. (2006), thereby enabling an independent assessment of relationships. In addition, RAG1 has been demonstrated to be phylogenetically informative for this level of relationship (e.g., López et al., 2004; Rüber et al., 2004; Holcroft, 2005; Sullivan et al., 2006; Chen et al., 2007; Mayden et al., 2007). Methods for DNA data collection followed standard procedures, as outlined in Conway et al. (2008), a previous study that utilized exon 3 of RAG1. The primers, RAG1F1 and RAG1R1, published in López et al. (2004) and the primer, R1-4061R, published in Chen et al. (2007) were used to amplify and sequence approximately 1500 bp of this loci. Sequences were deposited in GenBank (Table 1).

Bayesian analyses were conducted with the parallel version of MrBayes v.3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003; Altekar et al., 2004). Seven non-cypriniform outgroup taxa within the Ostariophysi were used as outgroups; *Chanos chanos* was designated as the most distant outgroup. Prior to the analysis, the sequence data were partitioned by codon position and MrModelTest v.2.2 (Nylander, 2004) and PAUP* (Swofford, 2002) were used to perform hierarchical likelihood ratio tests (hLRT) on each partition to determine the most appropriate model of nucleotide substitution. The best-fit model for all three codon positions of RAG1, in both the small and large data sets, was found to be GTR+I+Γ, which was applied during the MrBayes analyses. Two independent Bayesian searches were conducted for each data set (four total), each search ran for 1,000,000 generations, with 4 chains, sampling every 1,000 generations. The distribution of log likelihood scores was examined to determine stationarity and burn-in time for each search. In both searches involving the small data set, stationarity in log likelihood scores was observed after approximately 30,000–40,000 generations. Trees from the first 50,000 generations (51 trees) were discarded to ensure all burn-in trees were excluded. This left 950 trees from each search, which were combined to form a common pool of 1900 trees, these were then used to construct the 50% majority-rule consensus. In the two analyses of the large data set, stationarity was not observed until after 50,000–60,000 generations; to guarantee that all the trees examined were post-burn-in, the first 101 trees from each search (representing 100,000 generations) were discarded. The remaining 1800 trees were used to generate the 50% majority-rule consensus tree. Branch support for each clade was based on posterior probability values, indicated by the frequency of occurrence of each clade among the trees retained after the initial burn-in topologies were discarded.

2 Results and Discussion

As with any extremely diverse group, the phylogenetic relationships of Cypriniformes has had a troublesome history. In fact, some authors have described the relationships among species, genera, subfamilies and families as largely “chaotic” (Hubbs & Miller, 1977; Mayden, 1989; Mayden et al., 2007). Some taxa have been especially problematic in our ability to confidently decipher their sister-group relationships to develop a phylogenetically informative classification. The phylogenetic placement of *Gyrinocheilus* and *Tinca*, two of the most phylogenetically problematic genera in the order, and the naturalness of the traditionally recognized Balitoridae and Cobitidae (now recognized as two families, Botiidae and Cobitidae *sensu* Šlechtová et al., 2007), as well as the naturalness and phylogenetic relationships of the subfamilies of Catostomidae and Cyprinidae have all been difficult to seemingly intractable problems. Much of the difficulty with these systematic issues may, in part, owe its origin to the historic difficulties in character assessment, obtaining taxa for a global-wide Cypriniformes analysis, and the recent emergence of molecular analyses for the order (see Mayden et al., 2007). However, universal to all of the molecular studies for the order has been the examination of a limited number of taxa, in most cases less than 60–70 species and this degree of taxon sampling has only occurred in the last few years. Historically, even single morphological analyses of taxa in this order have been limited in the number of taxa and characters, with the notable exception of explicit phylogenetic studies by Sawada (1982), Mayden (1989), and Smith (1992). Consistency in the resolution of recent molecular trees for the traditionally recognized Cobitidae led Šlechtová et al. (2007) to elevate the subfamilies within Cobitidae (Botiinae and Cobitinae) to the family level, thus resolving the apparent polyphyly of the family. Their taxon sampling of the Balitoridae did not provide the diversity to identify the problems with this family identified herein. No analyses have focused on resolving the relationship of *Gyrinocheilus, Tinca,* or the subfamily naturalness and relationships within Cyprinidae.

Table 1 Taxa used in this study, with GenBank accession numbers for RAG1 sequences

Taxon	GenBank No.
Gonorynchiformes	
Chanidae	
Chanos chanos	AY430207
Gonorynchidae	
Gonorynchus greyi	EU409606
Siluriformes	
Callichthyidae	
Corydoras rabauti	Chen unpublished
Clariidae	
Clarias batrachus	DQ492521
Heteropneustidae	
Heteropneustes fossilis	DQ492522
Characiformes	
Alestidae	
Phenacogrammus interruptus	Chen unpublished
Characidae	
Chalceus macrolepidotus	EU409607
Cypriniformes	
Balitoridae	
Barbatula barbatula	EU711107
Barbatula toni	EU711133
Homaloptera leonardi	EU711130
Homaloptera parclitella	EU409610
Lefua echigonia	EF458305
Nemacheilus longicaudus	EU711124
Schistura balteata	EU711131
Sewellia lineolata	EU409609
Botiidae	
Botia striata	EU711109
Chromobotia macracantha	EU711137
Leptobotia mantschurica	EU711138
Leptobotia pellegrini	EU292683
Sinibotia superciliaris	EU711110
Catostomidae	
Catostomus commersonii	EU409612
Cycleptus elongatus	EU409613
Erimyzon oblongus	EU711117
Hypentelium nigricans	EU711134
Minytrema melanops	EU711135
Myxocyprinus asiaticus	EU711136
Thoburnia rhothoeca	EU711128
Cobitidae	
Acantopsis choirorhynchos	EU711139
Cobitis striata	Saitoh unpublished
Cobitis taenia	EU711113
Misgurnus anguillicaudatus	EU711122
Misgurnus nikolskyi	EU711140
Pangio oblonga	EU711141
Cyprinidae	
Acheilognathinae	
Acheilognathus typus	EU292688
Rhodeus atremius	EU711125
Rhodeus ocellatus	EU711142
Cultrinae	
Chanodichthys mongolicus	EU711145
Hemiculter lucidus	EU711119
Ischikauia steenackeri	EU292687
Cyprininae	
Barbonymus gonionotus	EU711146
Barbus barbus	EU711147
Barbus trimaculatus	EU711148
Capoeta capoeta	EU711111
Carassius auratus	DQ196520
Cyprinus carpio	AY787040
Garra orientalis	EU292684
Gymnocypris przewalskii	EU711149
Labeo batesii	EU711150
Labeo senegalensis	EU711151
Puntius ticto	EU711152
Puntius titteya	EU292685
Sawbwa resplendens	EU292686
Schizopyge curvifrons	EU711146
Gobioninae	
Abbottina rivularis	EU711102
Coreoleuciscus splendidus	EU711114
Gnathopogon elongatus	EU711153
Gobio gobio	EU292689
Hemibarbus labeo	EU711154
Pseudorasbora pumila	EU711155
Pungtungia herzi	EU711156
Romanogobio ciscaucasicus	EU409624
Sarcocheilichthys variegatus	EU711157
Leuciscinae	
Abramis brama	EU711103
Alburnoides bipunctatus	EU711104
Alburnus alburnus	EU711143
Aspius vorax	EU711106
Blicca bjoerkna	EU711108
Campostoma anomalum	EF452827
Clinostomus elongatus	EU711112
Couesius plumbeus	EU711115
Cyprinella lutrensis	EU711158
Erimystax dissimilis	EU711116
Exoglossum maxillingua	EU711118
Hemitremia flammea	EF452828
Hybognathus nuchalis	EU711120
Leucaspius delineatus	EU711121
Luxilus chrysocephalus	EF452829
Nocomis biguttatus	EF452830
Notemigonus crysoleucas	EF452831
Notropis atherinoides	EF452832
Notropis baileyi	EU292691
Opsopoeodus emiliae	EF452833
Pelecus cultratus	EU711144
Phoxinus percnurus	EU409627
Pimephales promelas	AY430210
Richardsonius balteatus	EF452835
Rutilus rutilus	EU711126
Scardinius erythrophthalmus	EU409628
Semotilus atromaculatus	EU409629
Tribolodon nakamurai	EU711159
Rasborinae	
Aphyocypris chinensis	EU292692
Aspidoparia morar	EU711105
Barilius bendelisis	EU292693
Boraras merah	EF452838
Chela dadiburjori	EU292694
Danio erythromicron	EU292698
Danio rerio	U71093
Danionella sp.	EF452841
Devario regina	EU292701
Esomus metallicus	EU292702
Horadandia atukorali	EU292703
Inlecypris auropurpurea	EU292708
Luciosoma setigerum	EU292704
Microrasbora rubescens	EU292706
Nicholsicypris normalis	EU711123
Opsaridium sp.	EF452846
Opsariichthys uncirostris	EF452847
Rasbora bankanensis	EU292709
Rasbora gracilis	EU292710
Sundadanio axelrodi	EU292711
Trigonostigma heteromorpha	EU711129
Zacco sieboldii	EU292713
Squaliobarbinae	
Ctenopharyngodon idella	EF178284
Tincinae	
Tinca tinca	EU711162
Xenocyprinae	
Xenocypris macrolepis	EU711160
Gyrinocheilidae	
Gyrinocheilus aymonieri	EU292682
Vaillantellidae	
Vaillantella maassi	EU711132

Unlike mitochondrial gene trees for the family, previous phylogenetic analyses at a higher level within Cypriniformes using nuclear genes are few and include only those by Mayden et. al. (2007), Šlechtová et al. (2007), Conway et al. (2008), and He et al. (2008b). The first study examined not only the phylogenetic placement of the model species *Danio rerio* but also evaluated some general relationships within the order. Šlechtová et al. (2007) assessed supraspecific relationships in the Cobitoidea. He et al. (2008b) focused on the subfamilies in Cyprinidae, exclusive of Psilorhynchinae, using the first intron of S7. Conway et al. (2008) reevaluated the classification of *Celestichthys margaritatus*, verifying it as a species of *Danio*, and also examined some higher relationships in the family Cyprinidae.

Phylogenetic relationships among the 53 targeted taxa within Cypriniformes, based on whole mitochondrial genomes (Fig. 1; Saitoh et al., 2006: fig. 2), identified a monophyletic Cypriniformes and two monophyletic superfamilies, Cobitoidea (Gyrinocheilidae, Catostomidae, Botiidae, Cobitidae, Balitoridae, Vaillantellidae) and Cyprinoidea (Cyprinidae with multiple natural and unnatural subfamilies). Nodal support for these relationships were very good overall, with posterior probabilities of 95–100%. Exceptions included relatively poor support for the superfamily Cobitoidea (63%) and similar support for one sister group hypothesis in Catotomidae (*Catostomus* + *Minytrema*) and some sister group relationships within the subfamily Gobioninae of Cyprinidae (*Gnathopogon* + (*Puntungia* + *Pseudorasbora*)). Within the Cobitoidea, *Gyrinocheilus* forms the sister group to Catostomidae; this Gyrinocheilidae + Catostomidae clade is sister to a loach clade consisting of monophyletic Botiidae, Cobitidae, and Balitoridae, along with the monotypic Vaillantellidae. *Vaillantella* is sister to a Cobitidae + Balitoridae clade, with Botiidae as the basal member of this loach clade. The sister group to the Cobitoidea is Cyprinidae, or the suborder Cyprinoidea. Within the family, the subfamily Cyprininae is sister to the remaining cyprinids. Saitoh et al. (2006) did not recover a monophyletic Rasborinae, instead they find support for a monophyletic Rasborinae *sensu stricto*, with the other putative rasborins found elsewhere in the tree. A monophyletic subfamily Acheilognathinae is recovered as the sister group to all remaining cyprinids (excluding Cyprininae and Rasborinae *sensu stricto*). A clade composed of Xenocyprinae, Cultrinae, and the remaining "rasborins" was recovered by Saitoh et al. (2006). The apical portion of the Cyprinidae is occupied by three subfamilies, Gobioninae and Leuciscinae, which are monophyletic, and the monotypic Tincinae. The subfamily Tincinae is sister to Leuciscinae, and that clade is sister to the Gobioninae.

With few exceptions, nodal support for both the smaller (49 cypriniform species) and larger (110 cypriniform species) analyses of taxa using only RAG1 sequences was generally high (90%–100%). Phylogenetic resolution among the 49 species examined herein for RAG1 (Fig. 2) had some notable differences from the topology of Saitoh et al. (2006) (Fig. 1). In terms of higher-level relationships, the RAG1-only tree (Fig. 2) did not find a monophyletic Cobitoidea, instead catostomids are found in a basal position as the sister group to a clade comprising all of the other cypriniforms, consequently Gyrinocheilidae is no longer the sister to Catostomidae, rather it is the sister group to a loach clade similar to the one seen in Fig. 1. Within the Catostomidae, relationships differed from the mitochondrial tree in that *Cycleptus* and *Myxocyprinus* did not form a clade. Rather, *Cycleptus* is the basal sister group in the family, and *Myxocyprinus* was sister to a clade wherein *Hypentelium* was sister to *Catostomus* plus *Minytrema*. Overall, the relationships within the loach clade (Botiidae, Vaillantellidae, Balitoridae, and Cobitidae) are roughly comparable to those seen in the mitogenome tree. The one major difference is that in the RAG1-only phylogeny, balitorids are not monophyletic, due to the position of *Homaloptera leonardi* (Fig. 2). The other major difference in the RAG1-only tree involves a radical rearrangement of the subfamilial relationships within the Cyprinidae. The subfamily Acheilognathinae is still monophyletic but instead of being the sister group to all cyprinids except Cyprininae and Rasborinae *sensu stricto*, it is sister to Gobioninae. Associated with that change is the relocation of *Tinca*, where it is now sister to a Leuciscinae + (Acheilognathinae + Gobioninae) clade. It should be noted that, although the tree obtained from our RAG1 analysis is not identical to the mitogenome topology presented by Saitoh et al. (2006), it is congruent with an alternate topology they obtained using the 12_nRT_n coding scheme excluding third codon positions (not shown; see Saitoh et al., 2006 for discussion of alternate topologies they recovered).

The relationships recovered by the analysis of the 110-taxa, RAG1-only data matrix (Fig. 3) resolve some of the conflicts between the two trees produced from examination of fewer taxa (Figs. 1 & 2). Beyond

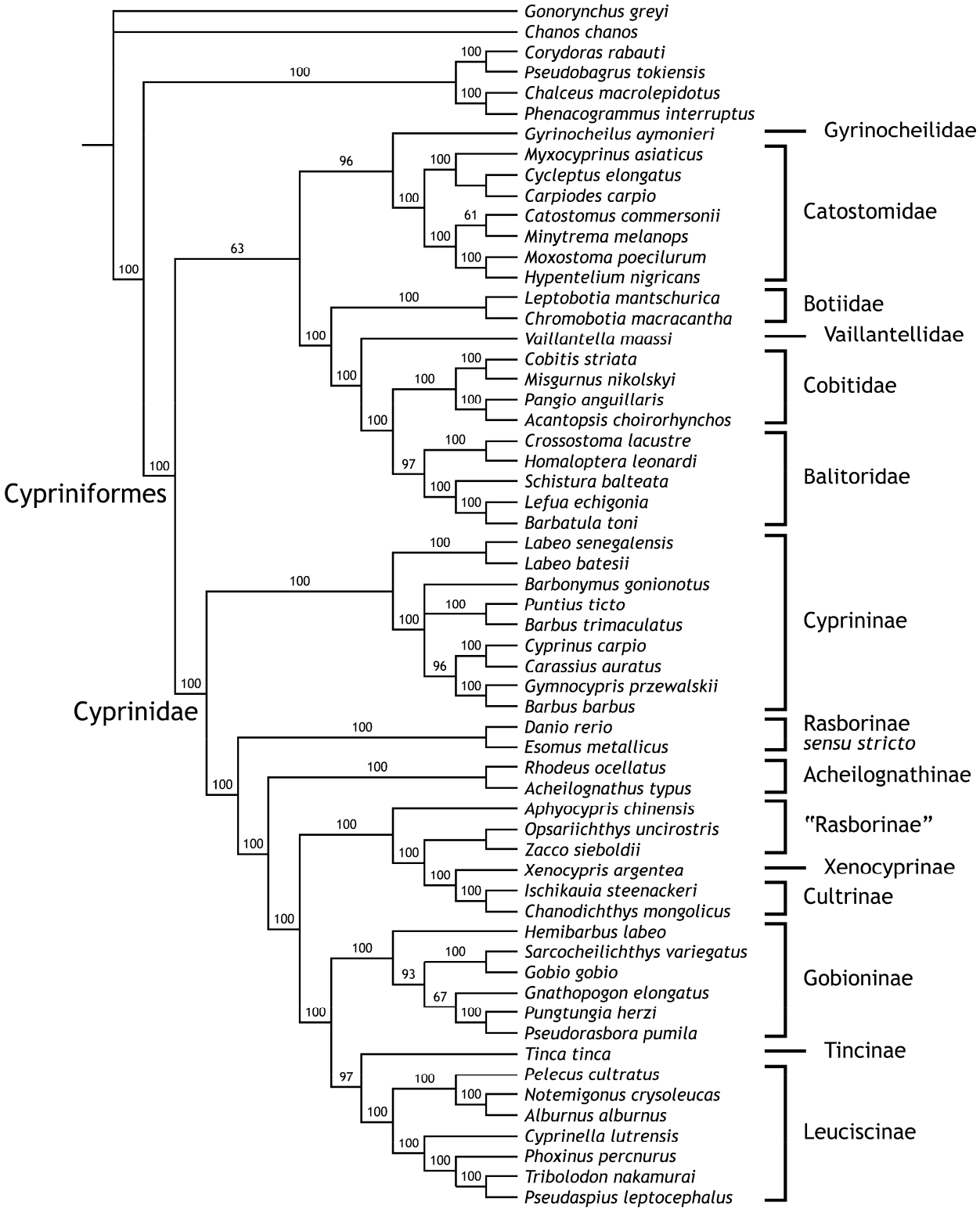

Fig. 1. A 50% majority rule consensus tree of 10,800 trees generated from Bayesian analysis of whole mitogenome sequence data from 53 cypriniform taxa, redrawn from Saitoh et al. (2006; fig. 2). Classification and family names are modified to reflect the taxonomy used herein (Figs. 2 & 3). Species names were drawn from information provided in Saitoh et al. (2006; table 2), with updates to reflect current nomenclature, following Eschmeyer (Catalog of Fishes, online version, updated 23 April 2008).

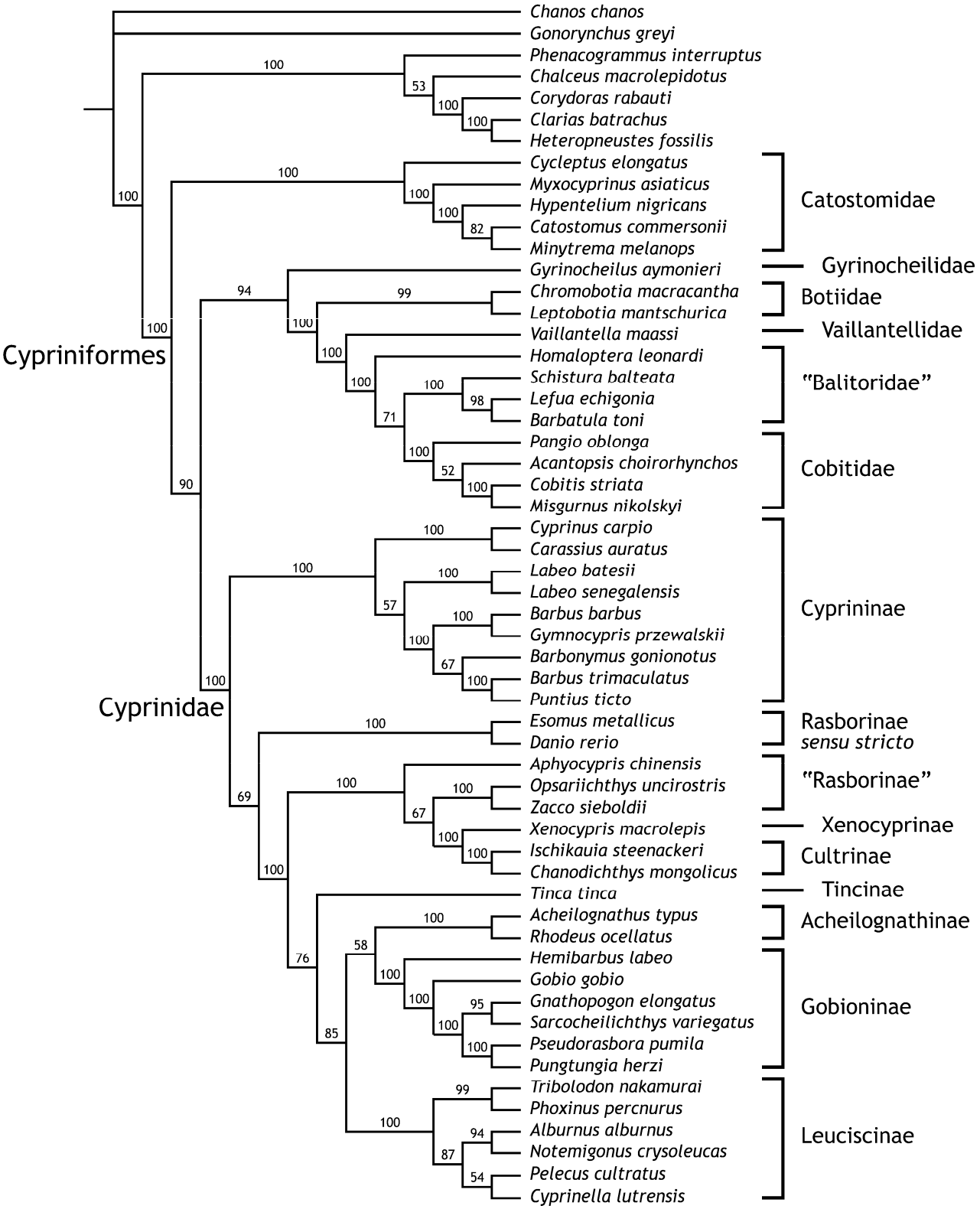

Fig. 2. A 50% majority rule consensus tree of 1,900 trees generated from Bayesian analysis of RAG1 sequence data collected from 49 cypriniform taxa, representing the same species as illustrated in Fig. 1 (Saitoh et al., 2006; fig. 2); when some species were not available, a congeneric species was used where possible. Bayesian posterior probabilities are displayed above each node.

the addition of taxa, the overall structure of the larger phylogeny is closely congruent with that observed in the 49-taxa, RAG1-only tree, in terms of familial and subfamilial relationships. In looking at the higher level relationships, the larger tree (Fig. 3) is congruent with the RAG1-only tree shown in Fig. 2. At the base of the tree, the superfamily Cobitoidea is not monophyletic, with Catostomidae as the most basal member of Cypriniformes, and Gyrinocheilidae is sister to a loach group. Catostomid relationships were completely consistent with those observed in the RAG1 phylogeny with 49 taxa except that *Thoburnia* and *Erimyzon* were included in the analysis; the former forms the sister group to *Hypentelium* and the latter is sister to *Minytrema*. Within Cyprinidae, the large tree agrees with the topology in Fig. 2: Acheilognathinae and Gobioninae as sister taxa, with Tincinae as the sister to a Leuciscinae + (Acheilognathinae + Gobioninae) clade. In addition, the monophyly of the Cyprininae, the Rasborinae *sensu stricto,* the Gobioninae, and the Leuciscinae is not challenged when more species from those groups are added. However, the Cultrinae becomes paraphyletic with respect to *Xenocypris macrolepis* in the larger analysis. The one major conflict between the mitogenome phylogeny and that using 49 taxa was the status of Balitoridae; a conflict that the incorporation of additional taxa resolved in favor of the Saitoh et al. (2006) tree. The inclusion of more balitorid species appears to solidify the monophyly of the Balitoridae (98% nodal support).

2.1 Relationships within Cobitoidea

Gyrinocheilus has been resolved as sister to a clade inclusive of Botiidae, Vaillantellidae, Balitoridae, and Cobitidae, sister to Catostomidae, or the basal sister group to all other Cypriniformes (He et al., 2008a). The former relationship is observed in both analyses herein for RAG1 and by Šlechtová et al. (2007), also based on analysis of RAG1 sequences. The sister relationship of Catostomidae and Gyrinocheilidae is observed in Saitoh et al. (2006) for 53 taxa and He et al. (2008a) for 17 ingroup taxa. One analysis in the latter study resolved *Gyrinocheilus* as the sister group to all other Cypriniformes. The variable relationships observed for the Cobitoidea cannot be resolved herein with the increase in taxon sampling as the results of all analyses appear to be partitioned on the basis of whether the character base is mitochondrial or nuclear. Future analyses of many more taxa for both mitochondrial and nuclear genes, and with an eye towards the relationships at this basal portion of the evolution of Cypriniformes, will be important in resolving this early diversification of the group.

2.2 Relationships within Catostomidae

Several hypotheses have been presented for relationships in Catostomidae, but all of the studies have had limited taxon sampling within the family and order. The mitochondrial phylogeny of Saitoh et al. (2006) identifies *Cycleptus* and *Myxocyprinus* as a basal monophyletic group, while that of Harris and Mayden (2001) identifies *Myxocyprinus* as the basal sister group in the family and *Cycleptus* sister to other taxa (excluding *Carpiodes* and *Ictiobus*, more basal in the tree). These studies are not, however, consistent in the resolution of *Carpiodes* and *Ictiobus*. Both RAG1 analyses identify *Cycleptus* as the basal sister group and *Myxocyprinus* as more closely related to other taxa (*Carpiodes* and *Ictiobus* not included in these analyses). The sister relationship of *Hypentelium* and *Thoburnia* and between *Minytrema* and *Erimyzon* is consistent with that of Harris and Mayden (2001) and Saitoh et al. (2006). However, the placement of the *Hypentelium* plus *Thoburnia* clade in the nuclear gene phylogeny is inconsistent with those analyses using fewer taxa (Saitoh et al., 2006; Harris & Mayden, 2001). While these differences could be the result of different resolutions based on alternative gene trees, we hypothesize that the placement of this clade is likely novel due to increasing taxon sampling, a hypothesis that should be tested using greater taxon sampling with the mitochondrial genes.

2.3 Relationships within Balitoridae

This group, as we recognize it today, was first proposed by Sawada (1982) based on morphological characters. In both instances, depending on the analysis, it has been resolved as either an unnatural or natural group. Previously, Tang et al. (2006) included a high number of species of the then recognized Cobitinae, Botiinae, Nemacheilinae, and Balitorinae and the relatively rapidly evolving mitochondrial genes cytochrome *b* and control region. In their study, the general relationships among these major groups were as observed herein for the larger sample of taxa (Fig. 3), except Tang et al. (2006) did not include the Vaillantellidae. The Cobitidae and Botiidae are separate monophyletic groups and Balitoridae is monophyletic and sister to the Cobitidae.

In other analyses with fewer taxa, the family Balitoridae is consistently resolved as a paraphyletic grade relative to the Cobitidae. This is not only observed in the present study but was also found by Tang et al. (2006), using cyt *b* sequence data, and later by He et al. (2008a) using whole mitochondrial gene sequences for only 17 species of Cypriniformes.

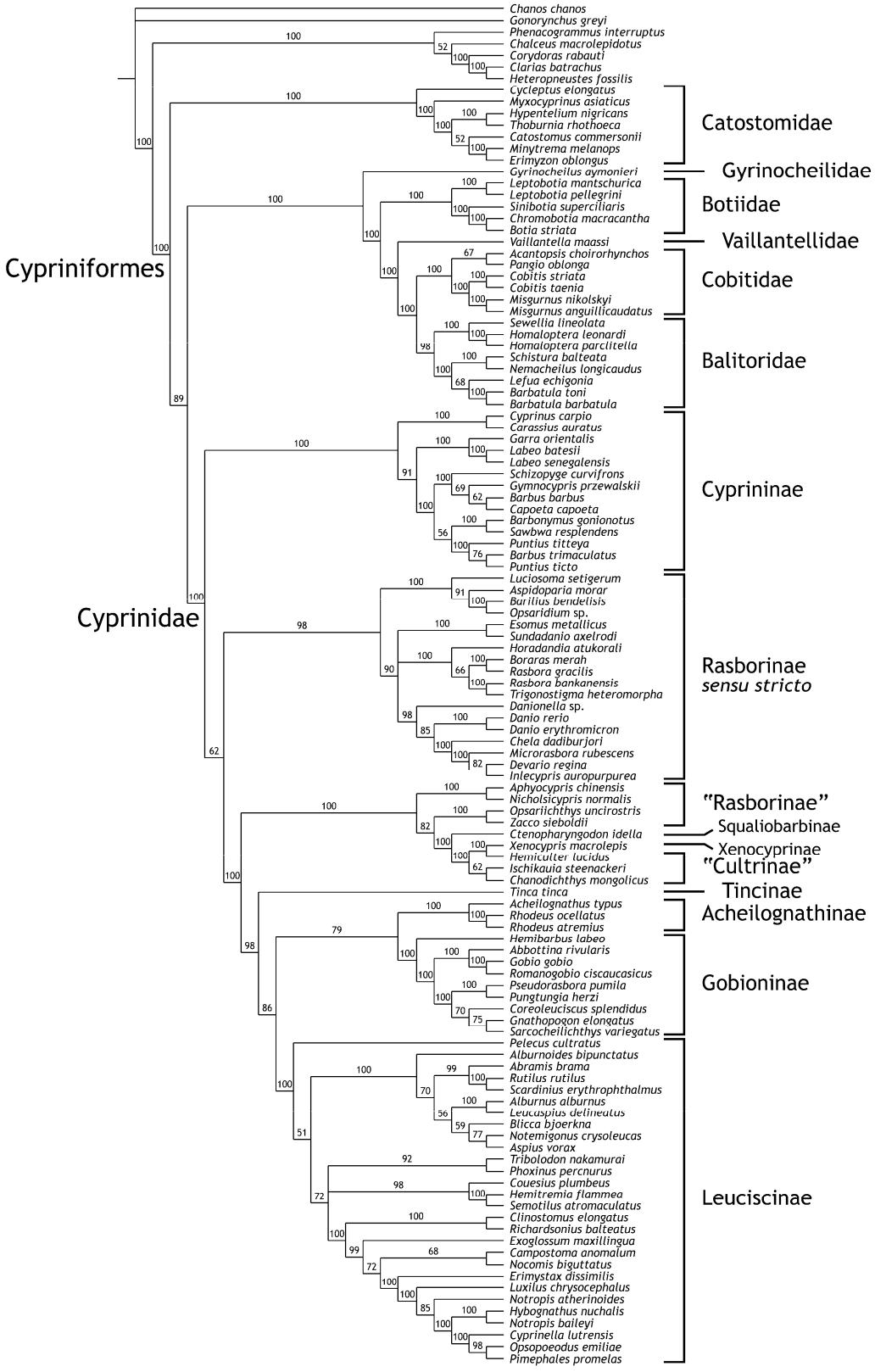

Fig. 3. A 50% majority rule consensus tree of 1,800 trees generated from Bayesian analysis of RAG1 sequence data collected from 110 cypriniform taxa, including all of the species shown in Fig. 2. Bayesian posterior probabilities are displayed above each node.

Interestingly, as observed for taxon sampling with the nuclear gene RAG1, with an increase in taxon sampling from these 17 species to 53 species in Saitoh et al. (2006) the family goes from being paraphyletic to monophyletic.

2.4 Monophyly of "Cobitidae" or Cobitidae and Botiidae

The unnaturalness of the historically conceived family Cobitidae was identified by Tang et al. (2006) and Saitoh et al. (2006), but taxonomic changes were not advocated in these analyses as caution was exercised in anticipation of more taxa and character data to provide further support that this resolution was not a result of long branch attraction in these highly morphologically divergent fishes. Šlechtová et al. (2007), however, recognized the Cobitidae and Botiidae (*sensu* Nalbant, 2002) and elevated the Nemacheilidae and Vaillantellidae based on RAG1 sequences. In this analysis the Cobitidae and Botiidae both resolve as monophyletic groups with either the smaller or larger taxon base. Further, both nuclear and mitochondrial gene analyses support these families as monophyletic groups. Although Šlechtová et al.'s (2007) Balitoridae and Nemacheilidae are recovered as reciprocally monophyletic groups in the large data set tree (Fig. 3), they also are recovered as sister groups, therefore we continue to recognize them as subfamilies of a monophyletic Balitoridae.

2.5 Relationships within Cyprinidae

The purported chaos regarding placement and relationships of species in the various subfamilies of Cyprinidae (Conway et al., in press) is not surprising. As currently conceived, this family is monophyletic (Cavender & Coburn, 1992; Saitoh et al., 2006; He et al., 2008b) but the vast majority of investigations of these species are pre-Hennigian revisionary and systematic studies, faunal works, and comparative taxonomic studies. There remains much diversity to be described and the current subfamilies' taxonomy rest largely on non-phylogenetic statements of inclusiveness that are essentially derived from phenetic similarity. Thus, it is expected that many of the forthcoming systematic studies of this family will result in changes in the taxonomy of the group but this is only because very few explicitly phylogenetic studies exist.

He et al. (2004) identified multiple, separate lineages for species referred to the Rasborinae, consistent with the polyphyletic origin of the subfamily observed by Saitoh et al. (2006), Mayden et al. (2007), Rüber et al. (2007), Conway et al. (2008) and herein for both analyses, regardless of taxon sample size (Figs. 2 & 3). The monophyly of Acheilognathinae and its sister relationship to the remaining apical cyprinids (excluding Cyprininae and Rasborinae *sensu stricto*), agrees with the results of Conway et al. (2008). The location of *Tinca* in the cypriniform tree of life is troublesome. However, the resolution of this lineage as the sister to the Leuciscinae + (Acheilognathinae + Gobioninae) clade is congruent with the Bayesian analyses derived from the nuclear intron S7 (He et al., 2008b: fig. 2C). The variable placement of *Tinca* in this instance may be related to the gene origin (mitochondrial versus nuclear), as well as the number and composition of taxa in the ingroup. The relative sister group relationships of taxa within the subfamilies Cyprininae, Gobioninae, and Leuciscinae are dependent on taxon sampling. In all three groups, the relationships observed in the larger taxon base are more consistent with previous studies (Cavender & Coburn, 1992; Simons & Mayden, 1997, 1998, 1999; Simons et al., 2003; He et al., 2008a, b) based on either morphological or molecular data.

Acknowledgements We thank Yin-Long Qiu and Zhiduan Chen and the Institute of Botany, Chinese Academy of Sciences, for extending an invitation to RLM for presenting this research at the Tree of Life symposium, June 2007 in Beijing, China. RLM wishes to thank David M. Hillis for great conversations regarding systematics and biodiversity, as well as the topic of focus in this paper, and urging us to publish this study. Finally, we thank the USA National Science Foundation for financial support for the Cypriniformes Tree of Life research (EF-0431326).

References

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

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20: 407–415.

Braun EL, Kimball RT. 2002. Examining basal avian divergences with mitochondrial sequences: Model complexity, taxon sampling, and sequence length. Systematic Biology 51: 614–625.

Bremer B, Jansen RK, 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). Systematic Biology 48: 413–435.

Bruno WJ, Halpern AL. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Molecular Biology and Evolution 16: 564–566.

Cavender TM, Coburn MM. 1992. Phylogenetic relationships of North American Cyprinidae. In: Mayden RL ed.

Systematics, historical ecology, and North American freshwater fishes. Palo Alto: Stanford University Press. 293–327.

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. Molecular Phylogenetics and Evolution 26: 262–288.

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 (Supplement B): 202–218.

Conway KW, Chen W-J, Mayden RL. 2008. The "Celestial Pearl danio" is a miniature *Danio* (s.s) (Ostariophysi: Cyprinidae): evidence from morphology and molecules. Zootaxa 1686: 1–28.

Conway KW, Hirt MV, Yang L, Mayden RL, Simons AM. Cypriniformes: Systematics and paleontology. In: Schultz H-P, Wilson MVH, Nelson JS eds. Origin and phylogenetic interrelationships of teleosts. Munich: Verlag Dr. Freidrich Feil. (in press)

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

DeBry RW. 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. Molecular Biology and Evolution 9: 537–551.

DeBry RW. 2005. The systematic component of phylogenetic error as a function of taxonomic sampling under parsimony. Systematic Biology 54: 432–440.

Dobzhansky T. 1973. Nothing in biology makes sense except in the light of evolution. American Biology Teacher 35: 125–129.

Eschmeyer WN, Ferraris CJ, Hoang MD, Long DJ. 1998. Species of fishes. In: Eschmeyer WN ed. Catalog of fishes. San Francisco: California Academy of Sciences. 25–1820. Online version, updated 23 April 2008.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27: 401–410.

Fitch WM, Beintema JJ. 1990. Correcting parsimonious trees for unseen nucleotide substitutions: The effect of dense branching as exemplified by ribonuclease. Molecular Biology and Evolution 7: 438–443.

Fitch WM, Bruschi M. 1987. The evolution of prokaryotic ferredoxins—with a general method correcting for unobserved substitutions in less branched lineages. Molecular Biology and Evolution 4: 381–394.

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

Gojobori T, Ishii K, Nei M. 1982. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. Journal of Molecular Evolution 18: 414–423.

Graham RL, Foulds LR. 1982. Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time. Mathematical Bioscience 60: 133–142.

Graybeal A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Systematic Biology 47: 9–17.

Groth JG, Barrowclough GF. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Molecular Phylogenetics and Evolution 12: 115–123.

Harris PH, Mayden RL. 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.

He S, Liu H, Chen Y, Kuwahara M, Nakajima T, Yang Z. 2004. Molecular phylogenetic relationships of Eastern Asian Cyprinidae (Pisces: Cypriniformes) inferred from cytochrome *b* sequences. Science in China (Series C Life Sciences) 47: 130–138.

He S, Gu X, Mayden RL, Chen W-J, Conway KW, Chen Y. 2008a. Phylogenetic position of the enigmatic genus *Psilorhynchus* (Ostariophysi: Cypriniformes): Evidence from the mitochondrial genome. Molecular Phylogenetics and Evolution 47: 419–425.

He S, Mayden RL, Wang X, Wang W, Tang KL, Chen W-J, Chen Y. 2008b. 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. Molecular Phylogenetics and Evolution 46: 818–829.

Heath TA, Hedtke SM, Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses. Journal of Systematics and Evolution 46: 239–257.

Hedtke SM, Townsend TM, Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Systematic Biology 55: 522–529.

Hendy MD, Penny D. 1989. A framework for the quantitative study of evolutionary trees. Systematic Zoology 38: 297–309.

Hennig W. 1950. Grundzüge einer Theorie der phylogenetischen Systematik. Berlin: Deutscher Zentralverlag.

Hennig W. 1966. Phylogenetic systematics (Davis DD, Zangerl R translation). Urbana: University of Illinois Press.

Hillis DM. 1996. Inferring complex phylogenies. Nature 383: 130–131.

Hillis DM. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Systematic Biology 47: 3–8.

Hillis DM, Pollock DD, McGuire JA, Zwickl DJ. 2003. Is sparse taxon sampling a problem for phylogenetic inference? Systematic Biology 52: 124–126.

Holcroft NI. 2005. A molecular analysis of the interrelationships of tetraodontiform fishes (Acanthomorpha: Tetraodontiformes). Molecular Phylogenetics and Evolution 34: 525–544.

Howes GJ. 1991. Systematics and biogeography: An overview. In: Winfield IJ, Nelson JS eds. Cyprinid fishes: systematics, biology and exploitation. London: Chapman and Hall.

Hubbs CL, Miller RR. 1977. Six distinctive cyprinid fish species referred to *Dionda* inhabiting segments of the Tampico embayment drainage of Mexico. Transactions of the San Diego Society of Natural History 28: 267–335.

Huelsenbeck JP. 1995. Performance of phylogenetic methods in

simulation. Systematic Biology 44: 17–48.
Huelsenbeck JP, Hillis DM. 1993. Success of phylogenetic methods in the 4-taxon case. Systematic Biology 42: 247–264.
Huelsenbeck JP, Lander KM. 2003. Frequent inconsistency of parsimony under a simple model of cladogenesis. Systematic Biology 52: 641–648.
Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754–755.
Hugall AF, Lee MSY. 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61: 2293–2307.
Inoue JG, Miya M, Tsukamoto K, Nishida M. 2001. A mitogenomic perspective on the basal teleostean phylogeny: Higher-level relationships with longer DNA sequences. Molecular Phylogenetics and Evolution 20: 275–285.
Johnson KP. 2001. Taxon sampling and the phylogenetic position of passeriformes: Evidence from 916 avian cytochrome *b* sequences. Systematic Biology 50: 128–136.
Krenz JG, Naylor GJ, Shaffer, HB, Janzen FJ. 2005. Molecular phylogenetics and evolution of turtles. Molecular Phylogenetics and Evolution 37: 178–191.
Kuhn T. 1962. The structure of scientific revolutions. Chicago: University of Chicago Press.
Lecointre G, Philippe H, Vân Lê HL, Le Guyader H. 1993. Species sampling has a major impact on phylogenetic inference. Molecular Phylogenetics and Evolution 2: 205–224.
Lin Y-H, McLenachan PA, Gore AR, Phillips MJ, Ota R, Hendy MD, Penny D. 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Molecular Biology and Evolution 19: 2060–2070.
López JA, Chen W-J, Ortí G. 2004. Esociform phylogeny. Copeia 2004: 449–464.
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.
Mayden RL. 1989. Phylogenetic studies of North American minnows, with emphasis on the genus *Cyprinella* (Teleostei: Cypriniformes). Miscellaneous Publication (University of Kansas Museum of Natural History) 80: 1–189.
Mayden RL, Tang KL, Conway KW, Freyhof J, Chamberlain SJ, Haskins M, Schneider L, Sudkamp M, Wood RM, Agnew M, Bufalino A, Sulaiman Z, Miya M, Saitoh K, He S. 2007. Phylogenetic relationships of *Danio* within the order Cypriniformes: A framework for comparative and evolutionary studies of a model species. Journal of Experimental Zoology B: Molecular and Developmental Evolution 308B: 642–654.
Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K., Shirai SM, Nishida M. 2003. Major patterns of higher teleostean phylogenies: A new perspective based on 100 complete mitochondrial DNA sequences. Molecular Phylogenetics and Evolution 26: 121–138.
Nalbant TT. 2002. Sixty million years of evolution. Part one: Family Botiidae (Pisces: Ostariophysi: Cobitoidea). Travaux du Muséum National d'Histoire Naturelle "Grigore Antipa" 44: 309–344.
Nelson JS. 2006. Fishes of the world. 4th ed. Hoboken: John Wiley and Sons, Inc.
Nylander JAA. 2004. MrModeltest v2.2. Uppsala: Evolutionary Biology Centre, Uppsala University.
Poe S. 2003. Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. Systematic Biology 52: 423–428.
Pollock DD, Zwickl DJ, McGuire JA, Hillis DM. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Systematic Biology 51: 664–671.
Rannala B, Huelsenbeck JP, Yang Z, Nielsen R. 1998. Taxon sampling and the accuracy of large phylogenies. Systematic Biology 47: 702–710.
Rokas A, Carroll SB. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Molecular Biology and Evolution 22: 1337–1344.
Rokas A, Krüger D, Carroll SB. 2005. Animal evolution and the molecular signature of radiations compressed in time. Science 310: 1933–1938.
Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425: 798–804.
Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.
Rosenberg MS, Kumar S. 2003. Taxon sampling, bioinformatics, and phylogenomics. Systematic Biology 52: 119–124.
Rüber L, Britz R, Kullander SO, Zardoya R. 2004. Evolutionary and biogeographic patterns of the Badidae (Teleostei: Perciformes) inferred from mitochondrial and nuclear DNA sequence data. Molecular Phylogenetics and Evolution 32: 1010–1022
Rüber L, Kottelat M, Tan HH, Ng PKL, Britz R. 2007. Evolution of miniaturization and the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate. BMC Evolutionary Biology 7: 38.
Saitoh K, Sado T, Mayden RL, 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.
San Mauro D, Gower DJ, Oommen OV, Wilkinson M, Zardoya R. 2004. Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. Molecular Phylogenetics and Evolution 33: 413–427.
Sawada Y. 1982. Phylogeny and zoogeography of the superfamily Cobitoidea (Cyprinoidei, Cypriniformes). Memoirs of the Faculty of Fisheries, Hokkaido University 28: 65–223.
Scotland RW, Olmstead RG, Bennett JR. 2003. Phylogeny reconstruction: The role of morphology. Systematic Biology 52: 539–548.
Simons AM, Mayden RL. 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 AM, Mayden RL. 1998. Phylogenetic relationships of the western North American phoxinins (Actinopterygii: Cyprinidae) as inferred from mitochondrial 12S and 16S RNA sequences. Molecular Phylogenetics and Evolution 9: 308–329.

Simons AM, Mayden RL. 1999. Phylogenetic relationships of North American cyprinids and assessment of homology of the open posterior myodome. Copeia 1999: 13–21.

Simons AM, Berendzen PB, Mayden RL. 2003. Molecular systematics of North American phoxinin genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Zoological Journal of the Linnean Society 139: 63–80.

Šlechtová V, Bohlen J, Tan HH. 2007. Families of Cobitoidea (Teleostei; Cypriniformes) as revealed from nuclear genetic data and the position of the mysterious genera *Barbucca, Psilorhynchus, Serpenticobitis* and *Vaillantella*. Molecular Phylogenetics and Evolution 44: 1358–1365.

Smith GR. 1992. Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America and Asia. In: Mayden RL ed. Systematics, historical ecology, and North American freshwater fishes. Palo Alto: Stanford University Press. 778–826.

Sorenson MD, Oneal E, García-Moreno J, Mindell DP. 2003. More taxa, more characters: The hoatzin problem is still unresolved. Molecular Biology and Evolution 20: 1484–1498.

Stamatakis A. 2006a. Phylogenetic models of rate heterogeneity: A high performance computing perspective. In: Parallel and Distributed Processing Symposium, 2006. IPDPS 2006, Rhodos, Greece, 25–29 April 2006.

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

Steppan SJ, Storz BL, Hoffmann RS. 2004. Nuclear DNA phylogeny of the squirrels (Mammalia: Rodentia) and the evolution of arboreality from c-myc and RAG1. Molecular Phylogenetics and Evolution 30: 703–719.

Sullivan JM, Swofford DL. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. Journal of Mammalian Evolution 4: 77–86.

Sullivan JP, Lundberg JG, Hardman M. 2006. A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using *rag1* and *rag2* nuclear gene sequences. Molecular Phylogenetics and Evolution 41: 636–662.

Susko E, Inagaki Y, Roger AJ. 2004. On inconsistency of the Neighbor-Joining, least squares, and minimum evolution estimation when substitution processes are incorrectly modeled. Molecular Biology and Evolution 21: 1629–1642.

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland: Sinauer Associates.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK eds. Molecular systematics, 2nd ed. Sunderland: Sinauer Associates. 407–514.

Tang Q, Liu H, Mayden RL, and Xiong B. 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). Molecular Phylogenetics and Evolution 39: 347–357.

Waddell PJ, Shelley S. 2003. Evaluating placental inter-ordinal phylogenies with novel sequences including RAG1, γ-fibrinogen, ND6, and mt-tRNA, plus MCMC-driven nucleotide, amino acid, and codon models. Molecular Phylogenetics and Evolution 28: 197–224.

Wegener A. 1915. Die Entstehung der Kontinente und Ozeane. Braunschweig: F. Vieweg and Sohn.

Wiens JJ. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Systematic Biology 54: 731–742.

Yang Z. 1994. Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. Systematic Biology 43: 329–342.

Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends in Ecology and Evolution 11: 367–372.

Zwickl DJ, Hillis DM. 2002. Increased taxon sampling greatly reduces phylogenetic error. Systematic Biology 51: 588–598.