



Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the world's largest clade of freshwater fishes: Further evidence from six nuclear genes

Wei-Jen Chen*, Richard L. Mayden

Department of Biology, Saint Louis University, 3507 Laclède Ave., St. Louis, MO 63103, USA

ARTICLE INFO

Article history:

Received 12 December 2008

Available online 21 January 2009

Keywords:

Cyprinidae

Cyprinoidea

Cypriniformes

*Psilorhynchus**Tinca**Leptobabrus*

Nuclear gene

Phylogenomics

Taxonomic chaos

1. Introduction

With over 210 genera and 2010 described species, the family Cyprinidae is currently the largest family of freshwater fishes (Nelson, 2006). Over the years, the Cyprinidae has been divided into different “groupings” for either taxonomic convenience or to represent presumed natural groups; usually these groupings have been recognized at or below the level of subfamily (Cavender and Coburn, 1992; Howes, 1991; Nelson, 2006). Howes (1991) recognized seven such subgroupings of the Cyprinidae, including the Alburninae, Cyprininae, Rasborinae, Cultrinae, Acheilognathinae, Tincinae, Leuciscinae, and Gobioninae. Cavender and Coburn (1992) recognized two, the Cyprininae and Leuciscinae, the former including those cyprinids referred to as barbinae, labeoninae and cyprininae, and the later including those referred to as tincinae, rasborinae, gobioninae, acheilognathinae, cultrinae, xenocyprinae, leuciscinae and phoxininae. The genus *Psilorhynchus* (a grouping of small, ventrally flattened fishes adapted for benthic life in fast flowing water) is either considered to be the sole member of the family Psilorhynchidae (Conway and Mayden, 2007; Nelson, 2006; Ramaswami, 1952) following Hora (1925) or as the sole member of the cyprinid subfamily Psilorhynchinae (Chen, 1981; Nelson, 1994).

Previous systematic analyses investigating monophyly and inter-relationships of the Cyprinidae have focused largely on morphology or mitochondrial gene/genome sequences. Chen et al. (1984) was the first such study to propose a “phylogenetic” hypotheses of cyprinid inter-relationships based on morphological data. Later Cavender and Coburn (1992) reanalyzed the data matrix of Chen et al. (1984), recovering a tree of equal length but of a different topology to that recovered by Chen et al. (1984) (Fig. 1B). Cavender and Coburn (1992) also proposed an alternative phylogeny for the Cyprinidae based on the analysis of their own 47 morphological characters (Fig. 1A). Despite these early morphological phylogenetic investigations of the Cyprinidae, some uncertainty regarding the basal lineage of cyprinids and the placement of the enigmatic genus *Tinca* remains (Fig. 1).

Recent systematic investigations of the Cyprinidae have utilized a molecular phylogenetic approach, with mitochondrial sequence data being most readily utilized (e.g., Cunha et al., 2002; Gilles et al., 2001; He et al., 2008a; Liu and Chen, 2003; Okazaki et al., 2001; Saitoh et al., 2006; Simons et al., 2003; Xiao et al., 2001; Zardoya and Doadrio, 1998). Although the taxonomic sampling of these studies was limited to certain subgroupings or to species from geographic regions of interests to the authors each of these investigations provided valuable insight into the evolution of these morphologically diverse fishes. Two studies presented their hypotheses with the samplings covering approximately all cyprinid subfamilies (Gilles et al., 2001; Liu and Chen, 2003). Both of these studies resolved *Tinca* as more closely related to members

* Corresponding author. Fax: +1 314 977 365.

E-mail address: wjchen.actinops@gmail.com (W.-J. Chen).

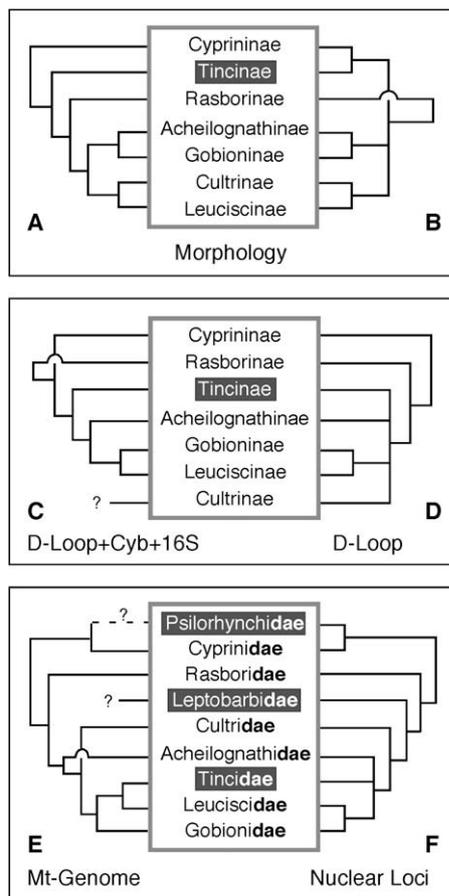


Fig. 1. Previous (A–E) and present (F) hypotheses depicting relationships of major lineages within the Cyprinidae. (A) Cavender and Coburn's (1992) hypothesis based on 47 morphological characters; (B) Most-parsimonious tree found by Cavender and Coburn (1992) based on morphological matrix of Chen et al. (1984); (C) Gilles et al.'s (2001) hypothesis based on 1374 aligned nucleotides from mt-DNA sequences of 16S, D-loop and cytochrome *b* genes; (D) Liu and Chen's (2003) hypothesis based on 1051 aligned nucleotides from mt-DNA sequences of D-loop; (E) Saitoh et al.'s (2006) hypothesis based on 14,563 aligned nucleotides from whole mt-genomic sequences; (F) Present hypothesis from this study based on 5733 aligned nucleotides from six nuclear genes with suggesting revised phylogenetic classification for the family (see: discussion). The three enigmatic taxa are highlighted. Leuciscinae(-dae) here includes leuciscines and phoxinines. Following results of previous studies of morphology and molecules, Cultrinae(-dae) should include cultrines, xenocyprines plus several other taxa suggested to be included to this group (see: discussion).

of the subfamilies Acheilognathinae, Gobioninae, and Cultrinae than to members of the Rasborinae or Cyprininae (Fig. 1C and D). More recently, using whole mitochondrial genomes, Saitoh et al. (2006) provided a robust phylogenetic hypothesis for the main cypriniform lineages for the first time. In their hypothesis, two reciprocal monophyletic groups were resolved: Cobitoidea and Cyprinoidea (or Cyprinidae). Their hypothetical relationships of main cyprinid clades were summarized in Fig. 1E. Even if a larger number of characters (e.g. complete mitochondrial genome data) was used, it has limitations to reaching a resolution of certain relationships within the family and some of critical and/or ambiguously classified taxa such as *Psilorhynchus* and *Leptobabrus* were missing in the analysis (Fig. 1). Moreover, the resulting hypotheses require further testing with more intensive taxonomic sampling and additional independent molecular markers.

Herein, we employ a multiple nuclear gene (or phylogenomic) approach with a diverse set of 49 cyprinid species to infer evolutionary relationships of the major clades within the Cyprinoidea. We placed particular emphasize on resolving the

phylogenetic position of the enigmatic genera *Psilorhynchus* (the stone carps), *Tinca* (the tench), and *Leptobabrus* (the mad barb or sultan fish) in relation to other cyprinid fishes. These enigmatic genera have received little attention from molecular systematists and have been difficult to place within the current cypriniform classification, likely because of morphological divergence. DNA sequence data were generated from six nuclear gene loci (RAG1, Rhodopsin, IRBP, EGR1, EGR2B, and EGR3). These gene markers have recently been shown to be phylogenetically informative in reconstructing the phylogenetic relationships of ray-finned fishes, particularly among fishes of the order Cypriniformes (Chen et al., 2008).

2. Materials and methods

2.1. DNA data collection

A total of 54 samples were included for investigation. The analytical dataset was composed of DNA sequences of 6 targeted nuclear loci obtained from 2 *Psilorhynchus* species, *Tinca tinca*, *Leptobarbus hoevenii*, 45 other diverse specimens of cyprinids from all recognized subfamily groups, and five outgroups from the superfamily Cobitoidea. Several sequences used in this study have been previously described in Mayden et al. (2008) and Chen et al. (2008). Methods for collecting new DNA data from the specimens and/or gene loci followed the procedures outlined in Chen et al. (2008). The GenBank accession numbers of corresponding gene sequences used in this study are listed in the Table 1.

2.2. Phylogenetic analysis

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 first matrix was composed of all available characters without employing a particular weighting scheme. 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 because of base composition biases across taxa (Chen et al., 2003; Lockhart et al., 1994), a second matrix (partial RY-coding matrix) was prepared according to the results obtained from absolute saturation tests (Philippe et al., 1994) and from χ^2 tests of base composition stationarity performed using PAUP*-version 4.0b10 (Swofford, 2002). As outlined in our previous study, no clear saturation plateau on substitutions in transitions at the third codon position of six nuclear genes used here was detected by comparing the sequences recovering all main lineages of cypriniform species (see Fig. 2 in Chen et al., 2008). However, the tests of base composition stationarity revealed that the Rhodopsin dataset exhibits significant base composition bias across taxa when analyzed using variable sites only and the sites at third codon position for the tests. Thus, we compiled an operational dataset in which the nucleotides A and G and the nucleotides T and C at the third codon position of Rhodopsin were converted into purine (R) and pyrimidine (Y), respectively.

Search for optimal ML trees and Bayesian analyses were performed by a high performance cluster computing facility (with 32 nodes) located at Saint Louis University. 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. Likelihood ratio tests (Goldman, 1993), as implemented in MrModeltest 2.2 (Nylander, 2004), were used to choose models for each gene coding position

Table 1
Cypriniform taxa included in this study and accession numbers of sequences in Genbank.

Family/subfamily	Taxon	GenBank accession no.					
		RAG1	RH	IRBP	EGR1	EGR2B	EGR3
Cobitoidea							
Balitoridae	<i>Sewellia lineolata</i>	EU409609	EU409635	EU409667	EU409699	EU409731	EU409763
Botiidae	<i>Leptobotia pellegrini</i>	EU292683	EU409640	EU409672	EU409704	EU409736	EU409768
Cobitidae	<i>Niwaella multifasciata</i>	EU409615	EU409642	EU409674	EU409706	EU409738	EU409770
Cyprinocheilidae	<i>Cyprinocheilus aymanieri</i>	EU292682	FJ197071	FJ197122	EU409727	EU409759	EU409791
Nemacheilidae	<i>Lefua costata</i>	EU409608	EU409634	EU409666	EU409698	EU409730	EU409762
Cyprinoidea							
Psilorhynchidae	<i>Psilorhynchus sucatio</i>	FJ531251	FJ531355	FJ531374	FJ531274	FJ531303	FJ531332
Psilorhynchidae	<i>Psilorhynchus homaloptera</i>	FJ531250	FJ531354		FJ531273	FJ531302	FJ531331
Cyprinidae							
Acheilognathinae	<i>Acheilognathus tabira</i>	EU409617	EU409644	EU409676	EU409708	EU409740	EU409772
Acheilognathinae	<i>Paracheilognathus himantegus</i>	EU409618	EU409645	EU409677	EU409709	EU409741	EU409773
Acheilognathinae	<i>Rhodeus ocellatus kurumeus</i>	EU711142	FJ197043	FJ197093	FJ531277	FJ531306	FJ531335
Cultrinae	<i>Ischikauia steenackeri</i>	EU292687	EU409648	EU409680	EU409712	EU409744	EU409776
Cultrinae	<i>Megalobrama amblycephala</i>	EU409620	EU409647	EU409679	EU409711	EU409743	EU409775
Cyprininae	<i>Acrossocheilus paradoxus</i>	FJ531245	FJ531342	FJ531362	FJ531255	FJ531284	FJ531313
Cyprininae	<i>Barbonymus gonionotus</i>	FJ531246	FJ531344	FJ531364	FJ531258	FJ531287	FJ531316
Cyprininae	<i>Barbus callipterus</i>	FJ531247	FJ531345	FJ531365	FJ531259	FJ531288	FJ531317
Cyprininae	<i>Garra spilota</i>	EU409621	EU409649	EU409681	EU409713	EU409745	EU409777
Cyprininae	<i>Gymnocypris przewalskii</i>	EU711149	FJ197051	FJ197102	FJ531265	FJ531294	FJ531323
Cyprininae	<i>Hampala macrolepidota</i>	EU409623	EU409651	EU409683	EU409715	EU409747	EU409779
Cyprininae	<i>Labeo chrysophekadion</i>	EU409622	EU409650	EU409682	EU409714	EU409746	EU409778
Cyprininae	<i>Puntius titteya</i>	EU292685	FJ531356	FJ531375	FJ531275	FJ531304	FJ531333
Gobioninae	<i>Biwia zezera</i>	EU409626	EU409654	EU409686	EU409718	EU409750	EU409782
Gobioninae	<i>Gobio gobio</i>	EU292689	FJ197056	FJ197107	FJ531264	FJ531293	FJ531322
Gobioninae	<i>Hemibarbus barbus</i>	EU711154	FJ197057	FJ197108	FJ531266	FJ531295	FJ531324
Gobioninae	<i>Romanogobio ciscaucasicus</i>	EU409624	EU409652	EU409684	EU409716	EU409748	EU409780
Gobioninae	<i>Sarcocheilichthys parvus</i>	EU409625	EU409653	EU409685	EU409717	EU409749	EU409781
Gobioninae	<i>Squalidus chankaensis</i>	FJ531252	FJ531358	FJ531377	FJ531278	FJ531307	FJ531336
Leuciscinae	<i>Notropis baileyi</i>	EU292691	EU409657	EU409689	EU409721	EU409753	EU409785
Leuciscinae	<i>Pelecus cultratus</i>	EU711144	FJ197045	FJ197095	FJ531272	FJ531301	FJ531330
Leuciscinae	<i>Phoxinus perenurus sachalinensis</i>	EU409627	EU409655	EU409687	EU409719	EU409751	EU409783
Leuciscinae	<i>Scardinius erythrophthalmus</i>	EU409628	EU409656	EU409688	EU409720	EU409752	EU409784
Leuciscinae	<i>Semotilus atromaculatus</i>	EU409629	EU409658	EU409690	EU409722	EU409754	EU409786
Rasboreinae	<i>Aphyocypris chinensis</i>	EU292692	FJ197066	FJ197117	FJ531256	FJ531285	FJ531314
Rasboreinae	<i>Aspidoparia morar</i>	EU711105	FJ531343	FJ531363	FJ531257	FJ531286	FJ531315
Rasboreinae	<i>Barilius bendelisis</i>	EU292693	FJ531346	FJ531366	FJ531260	FJ531289	FJ531318
Rasboreinae	<i>Danio albolineatus</i>	EU292696	EU409661	EU409693	EU409725	EU409757	EU409789
Rasboreinae	<i>Danio dangila</i>	EU292697	EU409660	EU409692	EU409724	EU409756	EU409788
Rasboreinae	<i>Danio rerio</i>	U71093	L11014	X85957	NM_131248	NM_130997	scaffold2320.1
Rasboreinae	<i>Danionella mirifica</i>	EU292700	FJ531347	FJ531367	FJ531261	FJ531290	FJ531319
Rasboreinae	<i>Devario regina</i>	EU292701	FJ531348	FJ531368	FJ531262	FJ531291	FJ531320
Rasboreinae	<i>Esomus longimanus</i>	FJ531248	FJ531349	FJ531369	FJ531263	FJ531292	FJ531321
Rasboreinae	<i>Horadandia atukorali</i>	EU292703	FJ531350	FJ531370	FJ531267	FJ531296	FJ531325
Rasboreinae	<i>Luciosoma setigerum</i>	EU292704	FJ531352	FJ531372	FJ531269	FJ531298	FJ531327
Rasboreinae	<i>Macrochirichthys macrochirus</i>	EU409630	EU409659	EU409691	EU409723	EU409755	EU409787
Rasboreinae	<i>Microrasbora kubotai</i>	EU292707	FJ531353	FJ531373	FJ531270	FJ531299	FJ531328
Rasboreinae	<i>Opsarichthys uncirostris</i>	FJ197126	FJ197068	FJ197119	FJ531271	FJ531300	FJ531329
Rasboreinae	<i>Rasbora bankanensis</i>	EU292709	FJ531357	FJ531376	FJ531276	FJ531305	FJ531334
Rasboreinae	<i>Rasbora steineri</i>	EU409631	EU409662	EU409694	EU409726	EU409758	EU409790
Rasboreinae	<i>Tanichthys albonubes</i>	FJ531253	FJ531359	FJ531378	FJ531279	FJ531308	FJ531337
Rasboreinae	<i>Trigonostigma heteromorpha</i>	EU292712	FJ531360	FJ531379	FJ531281	FJ531310	FJ531339
Rasboreinae	<i>Zacco sieboldii</i>	EU292713	FJ197069	FJ197120	FJ531283	FJ531312	FJ531341
Tincinae	<i>Tinca tinca</i>	EU711162	FJ197070	FJ197121	FJ531280	FJ531309	FJ531338
Unknown	<i>Leptobarbus hoevenii</i>	FJ531249	FJ531351	FJ531371	FJ531268	FJ531297	FJ531326
Unknown	<i>Paralabuca typus</i>	EU409619	EU409646	EU409678	EU409710	EU409742	EU409774
Unknown	<i>Yaoshanicus arcus</i>	FJ531254	FJ531361	FJ531380	FJ531282	FJ531311	FJ531340

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” (G + I), or “rates = gamma” (G), “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 ML search with the mixed model of nucleotide substitution we used a GTR + G + I model (with four discrete rate categories) for each partition because RAxML only provides GTR related models (GTR + G, GTR + G + I and GTR + CAT approximation) of rate heterogeneity for nucleotide data (Stamatakis, 2006). ML tree search was conducted by performing 100 distinct runs using the default algorithm of the program from complete random trees (-d option) as a starting tree for each run. The final tree was determined by a compar-

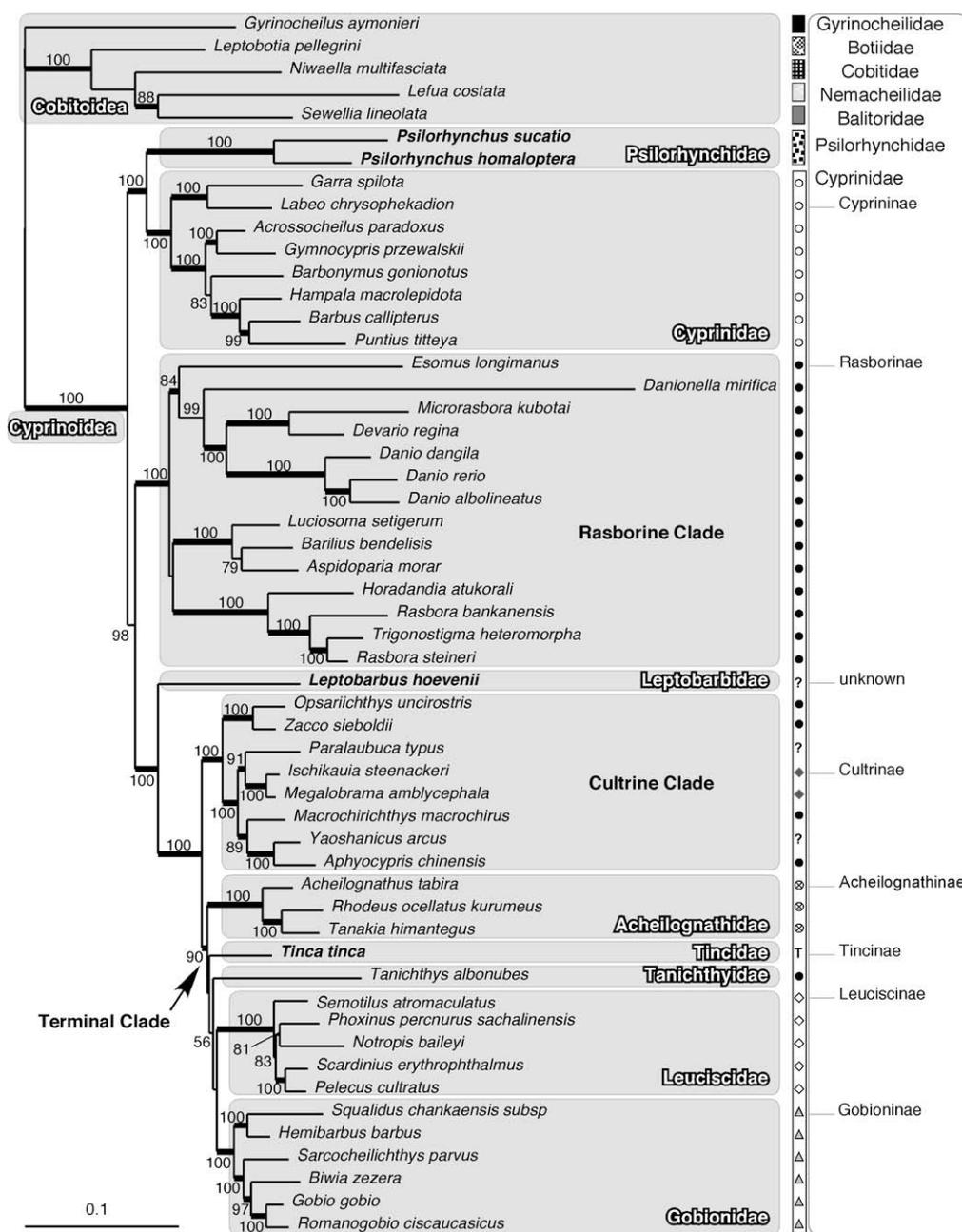


Fig. 2. Phylogenetic tree depicting relationships of the major clades resolved within the family Cyprinidae (or the superfamily Cyprinoidea). Relationships were obtained using partitioned ML analysis of 5733 aligned nucleotides from six nuclear gene loci. ML score of the tree is -53906.500279 . Branch lengths are proportional to inferred character substitutions under GTR + G + I model. Numbers on branches are ML bootstrap values; those 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%. The targeted taxa in this study, *Psilorhynchus*, *Tinca* and *Leptobarbus* are marked in bold. The bars and symbols on the right indicate traditional classification of taxa in Cypriniformes at family/subfamily level. A suggesting revised classification, based on robust molecular evidence from this study, is revealed by gray shadow rectangles on the topology.

ison of likelihood scores under GTR + G + 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 nonparametric 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 nonparametric bootstrap support (especially from MPBS) as potential upper and lower bounds of node robustness for our inferred phylogenetic trees (bold branches in Fig. 2). 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/.

3. Results and discussion

3.1. Characteristics of sequence data and inferred phylogenetic tree

A total of 5733 bp were aligned for the exon regions of six nuclear genes for 54 taxa (including five outgroups) sampled in this study. The length of aligned sequences from each locus was 1497 bp (RAG1), 819 bp (RH), 849 bp (IRBP), 846 bp (EGR1), 816 (EGR2B), and 906 (EGR3). No internal indels were found among aligned sequences for RAG1, RH, IRBP, and EGR2B datasets. A few indels needed to be introduced in adjusting sequences alignment of EGR1 and 3 genes, but the alignment can be unambiguously achieved followed by triplet codes for amino acids. Of 5733 nucleotides, 2497 were variable sites in which 1996 were parsimony informative. The second or partial RY-coding matrix presented 2407 variable sites in which 1901 were parsimony informative. Relationships of taxa derived from partitioned ML and Bayesian analyses of DNA sequences based on matrix 1 and 2 were nearly identical with slightly differences in relationships where nodal supports are weak; only the ML tree derived from the second (partial RY-coding) matrix is presented herein (Fig. 2). As shown, most of resulting clades were highly supported by partitioned MLBS, MPBS and *a posteriori* probabilities from partitioned BA (Fig. 2). Accordingly, 10 fully resolved clades or major lineages (as represented by gray shadow rectangles on the topology of Fig. 2) have emerged from the analyses in the present dataset.

3.2. Phylogenetic relationships of the major clades of cyprinid fishes

In all resulting phylogenies (all analyses), Cyprinidae is revealed as paraphyletic group with respect to the Psilorhynchidae. Within the Cyprinidae except for Rasborinae, the currently recognized cyprinid subfamilies were found to represent monophyletic groupings with strong nodal supports. As shown in ML tree with partial RY-coding analysis (Fig. 2), the Cyprininae and Psilorhynchidae are sister-groups to each other. These two clades together form the basal sister-group to the other cyprinid taxa shown in the tree.

Species currently placed within the Rasborinae appear to have five distinct origins among cyprinids. Among 19 rasborine species sampled in this study, 14 species group together in a robust clade or Rasborinae-2 clade *sensu* Conway et al. (2008). This rasborine clade can be subdivided into three strongly supported subgroups. The first subgroup includes species of *Luciosoma*, *Barilius* and *Aspidoparia*, which forms the sister-group to the subgroup containing species of *Rasbora*, *Trigonostiga* and *Horadandia* in the partial RY-coding ML analysis (Fig. 2). The remaining subgroup, which includes species of *Esomus*, *Danionella*, *Microrasbora*, *Devario* and *Danio*, forms the sister-group to a clade composed of two other subgroupings. In equal-weighting ML and BA analyses and partial RY-coding BA analysis, the first subgroup is resolved as the most basal lineage within the rasborine clade (not shown). This result corroborates to the findings from a recent study investigating inter-relationships among 31 rasborine taxa with RAG1 sequence data (Conway et al., 2008) but none of the studies with involved analyses resolve confidently the inter-relationships among these three mentioned rasborine subgroupings in terms of statistical nodal supports. Otherwise, most of intra-relationships within the subgroups are well resolved in this study (Fig. 2).

Four of five remaining “rasborine” taxa sampled in this study (*Opsariichthys*, *Zacco*, *Macrochirichthys*, and *Aphyocypris*) appear in three different placements in the tree and are more closely related to members of the Cultrinae and 2 other cyprinids with uncertain classification (*Yasoshanicus* and *Paralabuca*). We refer to this well-supported monophyletic group as the “cultrine” clade. These results, indicating a closer evolutionary affinity among certain rasborine species and cultrine species, are consistent with the findings of other previous phylogenetic analyses based on either mitochon-

drial or nuclear DNA sequence data (e.g. Saitoh et al., 2006; Conway et al., 2008; He et al., 2008b; Mayden et al., 2008).

The last rasborine species sampled in this study, *Tanichthys albonubes*, is nested within the terminal clade of the tree and separate from all other rasborines. Members from this clade represent many species endemic to Eurasia and North America from the cyprinid subfamilies Acheilognathinae, Gobioninae, Tincinae, and Leuciscinae (Fig. 2). Within the terminal clade, several subgroups were supported as monophyletic, notably at subfamily level. In all analyses, a sister-group relationship between the Leuciscinae and Gobioninae was recovered, but received only weak nodal support. This relationship is supported by previous molecular studies (Gilles et al., 2001; Liu and Chen, 2003) (Fig. 1), and in most of the resulting phylogenies based on varied analytical methods and datasets (four nuclear loci and whole mt-genomic data) in a recent study (Mayden et al., 2009).

Finally, another taxon of our interest, *Leptobarbus*, forms the sister-group to the cultrine clade plus the terminal clade of cyprinids described above. This relationship is strongly supported (100% for MLBP, BPBP, and *a posteriori* probabilities) (Fig. 2).

Overall, most of the phylogenetic relationships among cyprinid fishes presented here are well resolved using the DNA data from six nuclear loci (5733 bp). The results presented are largely congruent with the resulting cypriniform phylogeny using whole-mitogenome data (14,563 bp) (Saitoh et al., 2006). All of these steady molecular evidences currently established are challenging the morphological hypotheses and the classification of this group requires a further revision (see below: Section 3.4).

3.3. Systematics of *Tinca*, *Psilorhynchus*, and *Leptobarbus*

The systematic status of the genus *Tinca*, and *Psilorhynchus* is historically chaotic. As Howes (1991) stated, the inclusion of monotypic genus *Tinca* in any of the cyprinid subgroups is a taxonomic problem. For instance, two of the available morphological studies attempting to resolve the phylogenetic placement of *Tinca* among other cyprinids disagreed with each other (Cavender and Coburn, 1992; Chen et al., 1984) (Fig. 1A and B). Accordingly, erection of the subfamily Tincinae for *Tinca* became an optimal solution, yet its relationships remain uncertain. Early molecular hypotheses derived from mitochondrial sequence data rejected the morphological hypotheses and showed a closer evolutionary affinity of *Tinca* with acheilognathines, gobionines, leuciscines, and cultrines (Fig. 1C and D). Recent mt-genomic analysis identifies the phylogenetic position of the Tincinae, which should appear to be the sister-taxa to the Leuciscinae (Fig. 1E). Our resulting phylogeny, cannot further confirm this particular molecular hypothesis, as the corresponding nodes for those concerned relationships are weakly supported (Fig. 2). Nonetheless the hypothesis (as shown in Saitoh et al., 2006; Liu and Chen, 2003) (Fig. 1D and E) implicating that cultrine taxa have closely affinity to the Tincinae and to the remaining cyprinids (excluding cyprinines and “rasborines”) is less likely. Indeed, we resolve that *Tinca* is a member of the terminal clade of cyprinids (Fig. 1F; Fig. 2), the monophyly of which is highly supported.

Regarding the placement of the genus *Psilorhynchus* within the Cypriniformes, over the last two centuries this genus has been placed with different loach families (either Balitoridae or Cobitidae) or within the Cyprinidae (see Conway and Mayden, 2007; Šlechtová et al., 2007). The only morphological hypothesis applying phylogenetic analysis with relevant taxa from all families of the Cypriniformes showed that *Psilorhynchus* is the sister-taxa to a clade including cobitids and balitorids (Conway and Mayden 2007). Saitoh et al. (2006) did not include this taxa in their analysis. An alternative study using whole mt-genomic data with 17 representatives from the order Cypriniformes He et al. (2008a) identified *Psilorhynchus* and the Cyprininae as sister-group to each other. Our present study confirms this hypothesis.

Finally, prior to our study, no systematic study, inclusive of the genus *Leptobarbus*, had been conducted. *Leptobarbus* was included as member of the “Danioninae (possibly following Gosline, 1975) in the generic distribution list of cyprinids occurring in South East Asia by Rainboth (1991). In this study, we discover the systematic status of the genus *Leptobarbus* to be of intermediate position in the cyprinid tree (Fig. 2). Four described species are presented in the genus. Those are species native to South Asia from Thailand to Sumatra and Borneo. They may reach up to about 60 cm long. Interestingly, this huge taxon is the sister-group of an extremely diverse group comprising a large number of cyprinid species with highly diverse body shapes and sizes, occurring in different ecosystems.

3.4. Concluding remarks with taxonomy implication

According to the solid molecular evidence presented here, which supports the existence of ten monophyletic groups of cyprinid (or cyprinoid) fishes, we tentatively suggest the following revisions to the current cypriniform classification. Five of the subfamilies of the Cyprinidae (Cyprinae, Acheilognathinae, Tincinae, Leuciscinae, and Gobioninae) (Howes, 1991; Nelson, 2006) that have been previously recognized and widely believed to represent monophyletic groupings should be elevated from subfamily status to family status under the superfamily Cyprinoidea. The family group name Psilorhynchidae, which has been previously accepted by numerous authors (Conway and Mayden, 2007; Nelson, 2006; Ramaswami, 1952) should be retained. Erection of new cyprinoid families (Leptobarbidae and Tanichthyidae) for two distinct lineages revealed from our analyses containing the species from *Leptobarbus* and *Tanichthys*, respectively is recommended. We would also need to give the relevant family names for the clades discovered in this study. There are two clades (rasborine and cultrine clades) containing, respectively, major rasborine species and cultrine species plus their putative allies such as *Opsariichthys*, *Zacco*, *Paralabuca*, *Macrochirichthys*, *Yaoshanicus*, *Aphyocypris*, taxa from “Xenocyprinae”, and taxa from “Squaliobarbinae” that are identified in this study (Fig. 2) and in some other molecular studies (He et al., 2008b; Liu and Chen, 2003; Mayden et al., 2008; Saitoh et al., 2006). Finally, ten families herein are suggested to be recognized for the Cyprinoidea with respect to seven families for its reciprocal superfamily Cobitoidea (Šlechtová et al., 2007).

Acknowledgments

We greatly thank Kevin L. Conway for providing valuable comments for this work and for manuscript-improvement. We thank Maski Miya, Tetsuya Sado, and Shunping He for the specimens. This research is part of an ongoing international Tree of Life initiative on the order Cypriniformes USA NSF awards EF 0431326.

References

- Cavender, T.M., Coburn, M., 1992. Phylogenetic relationships of North American Cyprinidae. In: Mayden, R.L. (Ed.), Systematics, Historical Ecology and North American Freshwater Fishes. Stanford University Press, Stanford, pp. 328–378.
- Chen, W.-J., Bonillo, C., Lecointre, G., 2003. Repeatability of clades as 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., Miya, M., Saitoh, K., Mayden, R.L., 2008. Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: the order Cypriniformes (Ostariophysi) as a case study. Gene 423, 125–134.
- Chen, X., Yue, P., Lin, R., 1984. Major groups within the family Cyprinidae and their phylogenetic relationships. Act. Zoo. Sinica 9, 424–440.
- Chen, Y., 1981. Investigation of the systematic position of *Psilorhynchus* (Cyprinoidei, Pisces). Acta Hydrobio. Sinica 7, 371–376.
- Conway, K.W., Mayden, R.L., 2007. The gill-arches of *Psilorhynchus* (Ostariophysi: Psilorhynchidae). Copeia 2007, 267–280.
- Conway, K.W., Chen, W.-J., Mayden, R.L., 2008. The celestial pearl danio is a miniature *Danio* (s.s.) (Ostariophysi: Cyprinidae): evidence from morphology and molecules. Zootaxa 1686, 1–28.
- 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.
- Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
- Gilles, A., Lecointre, G., Miquelis, A., Loerstcher, M., Chappaz, R., Brun, G., 2001. Partial combination applied to phylogeny of European cyprinids using the mitochondrial control region. Mol. Phylogenet. Evol. 19, 22–33.
- Goldman, N., 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36, 182–198.
- Gosline, W.A., 1975. The cyprinid dermosphenotic and the subfamily Rasborinae. Occ. Pap. Mus. Zool., Univ. Mich 673, 1–13.
- He, S., Gu, X., Mayden, R.L., Chen, W.-J., Conway, K.W., Chen, Y., 2008a. Phylogenetic position of the enigmatic genus *Psilorhynchus* (Ostariophysi: Cypriniformes): evidence from the mitochondrial genome. Mol. Phylogenet. Evol. 47, 419–425.
- He, S., Mayden, R.L., Wang, X., Wang, W., Tang, K.L., 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. Mol. Phylogenet. Evol. 46, 818–829.
- Hora, S.L., 1925. Notes of fishes in the Indian Museum. 12. The systematic position of the cyprinoid genus *Psilorhynchus* McClelland. Rec. Indian Mus. 27, 457–460.
- Howes, G.J., 1991. Systematics and biogeography: an overview. In: Winfield, I.J., Nelson, J.S. (Eds.), Cyprinid Fishes Systematics, Biology and Exploitation. Chapman and Hall, London, pp. 1–33.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES. Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
- Liu, H., Chen, Y., 2003. Phylogeny of the East Asian cyprinids inferred from sequences of mitochondrial DNA control region. Can. J. Zool. 81, 1938–1946.
- 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.
- Mayden, R.L., Chen, W.-J., Bart, H.L., Doosey, M.H., Simons, A.M., Tang, K.L., Wood, R.M., Agnew, M.K., Yang, L., Hirt, M.V., Clements, M.D., Saitoh, K., Sado, T., Miya, M., Nishida, M., 2009. Reconstructing the phylogenetic relationships of the Earth's most diverse clade of freshwater fishes – order Cypriniformes (Actinopterygii: Ostariophysi): a case study using multiple nuclear loci and the mitochondrial genome. Mol. Phylogenet. Evol. 51, 500–514.
- 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., Wang, X., Saitoh, K., He, S., Liu, H., Chen, 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.
- Nelson, J.S., 1994. Fishes of the World, third ed. John Wiley & Sons, Inc., New York.
- Nelson, J.S., 2006. Fishes of the World, fourth ed. John Wiley & Sons, Inc., Hoboken, NJ.
- Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- Okazaki, M., Naruse, K., Shima, A., Arai, R., 2001. Phylogenetic relationships of bitterlings based on mitochondrial 12S ribosomal DNA sequences. J. Fish Biol. 58, 89–106.
- Philippe, H., Sorhannus, U., Baroin, A., Perasso, R., Gasse, F., Adoutte, A., 1994. Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. Mol. Phylogenet. Evol. 7, 247–265.
- Rainboth, W.J., 1991. Cyprinids of South East Asia. In: Winfield, I.J., Nelson, J.S. (Eds.), Cyprinid Fishes Systematics, Biology and Exploitation. Chapman and Hall, London, pp. 156–210.
- Ramaswami, L.S., 1952. Skeleton of cyprinoid fishes in relation to phylogenetic studies. II. The systematic position of *Psilorhynchus* McClelland. Proc. Nat. Inst. Sci. India 18, 141–150.
- Saitoh, K., Sado, T., Mayden, R.L., Hanzawa, N., Nakamura, K., Nishida, M., Miya, M., 2006. Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): the first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. J. Mol. Evol. 63, 826–841.
- Simons, A.M., Berentzen, P., Mayden, R.L., 2003. Molecular systematics of North American phoxinini genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Zool. J. Linn. Soc. 139, 63–80.
- 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 *Barbus*, *Psilorhynchus*, *Serpenticobitis* and *Vaillantella*. Mol. Phylogenet. Evol. 44, 1358–1365.
- 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.
- Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Xiao, W., Zhang, Y., Liu, H., 2001. Molecular systematics of Xenocyprinae (teleostei: cyprinidae): taxonomy, biogeography, and coevolution of a special group restricted in East Asia. Mol. Phylogenet. Evol. 18, 163–173.
- Zardoya, R., Doadrio, I., 1998. Phylogenetic relationships of Iberian cyprinids: systematic and biogeographical implications. Proc. Biol. Sci. 265, 1365–1372.