



Journal of Fish Biology (2009) **75**, 2197–2208

doi:10.1111/j.1095-8649.2009.02398.x, available online at www.interscience.wiley.com

Molecular phylogeny of the Cobitoidea (Teleostei: Cypriniformes) revisited: position of enigmatic loach *Ellopostoma* resolved with six nuclear genes

W.-J. CHEN*†‡, V. LHEKNIM§ AND R. L. MAYDEN*

**Department of Biology, Saint Louis University, 3507 Laclede Ave, St. Louis, MO 63103, U.S.A.* and §*Department of Biology, Prince of Songkla University, P.O. Box 3, Ko Hong, Songkhla 90112, Thailand*

(Received 10 March 2009, Accepted 16 July 2009)

Molecular variation in six nuclear genes provides substantive phylogenetic evidence for the recognition of a new cypriniform family, the Ellopostomatidae, to include the enigmatic Southern Asia loach genus *Ellopostoma*. The current six loach families form a monophyletic group, with the Nemacheilidae as the sister group to Ellopostomatidae; Vaillantellidae forms the sister group to all families exclusive of Botiidae. While the superfamily Cobitoidea includes eight families, the monophyly of this large clade within the Cypriniformes remains a vexing problem despite extensive molecular analyses and is in need of further investigation.

© 2009 The Authors

Journal compilation © 2009 The Fisheries Society of the British Isles

Key words: Cobitoidea; Cypriniformes; *Ellopostoma*; loaches; nuclear gene phylogeny.

INTRODUCTION

The order Cypriniformes contains many culturally, economically (*e.g.* carps) and scientifically important species [*e.g.* the model organism species *Danio rerio* (Hamilton), zebrafish], is currently the largest monophyletic group of freshwater fishes, with over 400 genera and *c.* 5000 species (described and undescribed) and is native to Asia, Europe, Africa and North America (Nelson, 2006). Concomitant with this extensive radiation has been the evolution of a great diversity in morphology, ecology, physiology, distribution and other life-history aspects. This notable diversification is interesting from an evolutionary perspective and has resulted in a proliferation of research focused on these fishes. However, evolutionary studies must be grounded in robust investigations of the genealogical relationships among taxa. Thus, research on the systematic and evolutionary studies of Cypriniformes has potential for complementary and valuable information in comparative biology, conservation and aquaculture (Mabee *et al.*, 2007; Mayden *et al.*, 2007; Schilling & Webb, 2007).

†Author to whom correspondence should be addressed. Tel.: +886 (0)23366 1630; fax: +886 (0)22362 6092; email: wjchen.actinops@gmail.com

‡Present address: Institute of Oceanography, National Taiwan University, No. 1 Sec. 4 Roosevelt Rd., Taipei 10617, Taiwan

Cypriniformes is currently divided into two superfamilies (Nelson, 2006) following Siebert (1987): the Cobitoidea (loaches and allies) and the Cyprinoidea (carps, minnows and allies). For nearly a century, this clade of fishes has posed several taxonomic challenges to ichthyologists and has been one of the most problematic groups of fishes in the systematics of ray-finned fishes, Actinopterygii. Classification within the Cobitoidea has varied, following the conclusions of various authors studying the evolutionary affinities among taxa based on morphology (see Table I of Šlechtová *et al.*, 2007). In general, three distinct groupings within the superfamily have been recognized, including the Gyриноcheilidae (algae eaters), Catostomidae (suckers) and the most diverse group, the loaches (Cobitidae and Balitoridae) (Sawada, 1982; Nelson, 2006). With advances in the acquisition and analysis of molecular data, recent phylogenetic results from both mitochondrial (mt) (Saitoh *et al.*, 2006; Tang *et al.*, 2006) and nuclear genomes (Šlechtová *et al.*, 2007; Mayden *et al.*, 2008; Mayden *et al.*, 2009) have been used to a greater extent in attempts to resolve the classification of monophyletic groups within the Cobitoidea. Evidence from these studies converge toward supporting seven cobitoid families: Gyриноcheilidae, Catostomidae, Cobitidae, Botiidae (formerly included in Cobitidae), Balitoridae, Nemacheilidae (formerly included in Balitoridae) and Vaillantellidae [newly erected by Šlechtová *et al.* (2007)] (Mayden *et al.*, 2009). However, despite these efforts, the resolution of the higher level, sister-group relationships among these families remains largely unresolved because of the use of either a limited number of genes or taxa. Using whole mitochondrial genomes for 53 in-groups and six out-groups, Saitoh *et al.* (2006) provided a phylogenetic hypothesis for the major cypriniform lineages, for the first time. However, some relationships within the Cobitoidea (*e.g.* sister-group relationship for Vaillantellidae) in this study were inconsistent, based on various analytical methods or data matrices (see also Mayden *et al.*, 2009). Furthermore, the hypothesis, based on the mt genomes, depicting the sister-group relationship between Gyриноcheilidae and Catostomidae (see also He *et al.*, 2008), has not been supported by later analyses, based on nuclear genes (Šlechtová *et al.*, 2007; Mayden *et al.*, 2008; Mayden *et al.*, 2009).

One of the most enigmatic of the cobitoid taxa is the genus *Ellopostoma* Vaillant, a taxon that till now has been absent in all molecular analyses of the Cypriniformes. The genus currently includes two described species. *Ellopostoma megalomycter* (Vaillant) was the sole member of the genus since it was first discovered. This species is native to Southeast Asia from peninsular Malaysia to western Borneo. A second species, *Ellopostoma mystax* Tan & Lim (Tan & Lim, 2002), was recently described and is thought to be endemic to peninsular Thailand. Species of *Ellopostoma* are morphologically adapted to a benthic lifestyle and inhabit moderate to swift-flowing rivers. They are moderately elongate, have enlarged nostrils and eyes, possess small scales and have a peculiar inferior mouth, features that together characterize their bizarre appearance among cypriniform fishes (Roberts, 1972) (Fig. 1). *Ellopostoma* was originally assigned to the family Cobitidae (Vaillant, 1902). However, at that time, members of the unsettled Cobitidae could be referred to any loach-like lineage known (Šlechtová *et al.*, 2007). Since the discovery of this taxon, several morphological studies have had difficulty placing the taxon within existing cypriniform classifications because of its unique morphology (Roberts, 1972; Kotelat, 1989; Roberts, 1989; Banarescu & Nalbant, 1995; Tan & Lim, 2002). The

TABLE I. Gene partitions assigned for the analyses and best models of nucleotide substitution chosen for each partition in the BA analysis

No.	RAG1			Rhodopsin			IRBP		
	1st pos.	2nd pos.	3rd pos.	1st pos.	2nd pos.	3rd pos.	1st pos.	2nd pos.	3rd pos.
1	GTR+G+I	GTR+G+I	GTR+G+I	4	5	6	7	8	9
Model	GTR+G+I	GTR+G+I	GTR+G+I	GTR+G+I	GTR+G+I	GTR+G+I	GTR+G+I	GTR+G+I	GTR+G+I
EGR1									
EGR2B									
EGR3									
10	HKY+G+I	GTR+G+I	GTR+G	13	14	15	16	17	18
Model	HKY+G+I	GTR+G+I	GTR+G	GTR+G+I	GTR+I	GTR+G	GTR+G+I	GTR+I	GTR+G

Note: pos., codon position of protein-coding genes.



FIG. 1. *Ellopostoma mystax*, adult male collected from Khlong Sok near Phanom District, Thailand. Photograph by Vachira Lheknim & Suebpong Sa-nguansil.

current taxonomy of the genus is cited as Balitoridae: Nemacheilinae, following references in Eschmeyer's *Catalogue of Fishes* (Eschmeyer & Fricke, 2008), but with comments for the need of revision.

In the present study, an expansive set of nuclear DNA sequences were used to test (1) current molecular hypotheses of the high-level systematics of the Cobitoidea and (2) the phylogenetic position of the enigmatic loach *E. mystax* (Tan & Lim, 2002) in relation to other cypriniform fishes.

MATERIALS AND METHODS

This study included 49 fish. The analysis included DNA sequences from six nuclear loci in one *E. mystax* (captured in Tapi River Basin, Surat Thani Province, south Thailand: 8° 50' N; 99° 19' E), 44 specimens of other diverse cypriniform lineages, including 24 species from the Cobitoidea, 20 species from the Cyprinoidea and four outgroups from other ostariophysans. Several sequences previously appeared in Chen *et al.* (2008) and Mayden *et al.* (2008, 2009). For collecting new DNA data from the specimens or loci, or both, methods of Chen *et al.* (2008) were followed. GenBank accession numbers of sequences used in this study appear in the Appendix.

Phylogenetic analyses were based on a partitioned maximum likelihood (ML) method and partitioned Bayesian approach (BA) for two character matrices as implemented in the parallel version of RAxML 7.0.4 (Stamatakis, 2006) and MrBayes 3.1.1 (Huelsenbeck & Ronquist, 2001), respectively. The first matrix consisted of all available characters without a weighting scheme. As phylogenetic analyses of protein-coding genes can be biased by homoplasies at third codon positions because of multiple substitutions in transitions (Saitoh *et al.*, 2006) or because of base composition biases across taxa, or both (Lockhart *et al.*, 1994; Chen *et al.*, 2003), a second matrix (partial RY-coding matrix) was prepared according to the results from absolute saturation tests (Philippe *et al.*, 1994) and from χ^2 tests of base composition stationarity performed with PAUP* 4.0b10 (Swofford, 2002). As in Chen *et al.* (2008), no clear saturation plateau on substitutions in transitions at the third codon position of the six nuclear genes used here appeared in sequences from all the major cypriniform lineages (see Fig. 2 in Chen *et al.*, 2008). However, tests of base composition revealed that variable sites and sites at a third codon position in RAG1, rhodopsin, EGR2B and EGR3 sequences exhibit significant base composition bias across taxa. Thus, an additional dataset was constructed in which the nucleotides A and G and the nucleotides T and C at the third codon position of these four genes 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 (20 nodes) at Saint Louis University. A mixed model analysis was used, which allows the independent estimation of individual models of nucleotide substitution for each partition. For the targeted nuclear loci, 18 partitions were assigned for all of

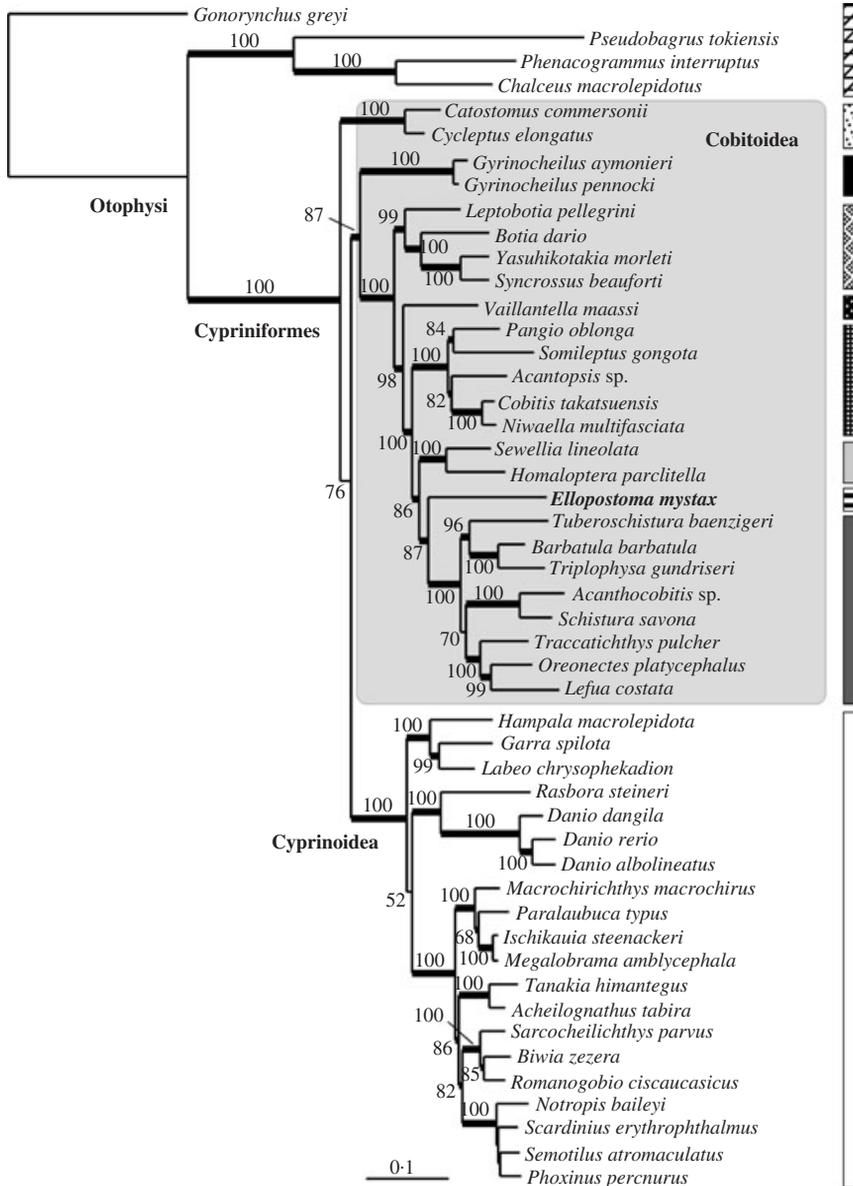


Fig. 2. Phylogenetic tree depicting relationships among taxa of the Cobitoidea and cypriniform allies. Tree was constructed using partitioned ML analysis of 5733 aligned nucleotides from six nuclear loci in 18 partitions assigned with respect to the gene and the codon positions. ML score of the tree is -44120.602658. Branch lengths are proportional to inferred character substitutions under the 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 . The targeted taxon in this study, *Ellopostoma*, is also marked in bold. Classification follows Šlechtová *et al.* (2007). A new family, Ellopostomatidae, including species from *Ellopostoma* is recommended. See Chen & Mayden (2009) for a detailed revision of the classification and molecular systematics of Cyprinoidea based on this same set of genes. [Pattern 1], Balitoridae; [Pattern 2], Botiidae; [Pattern 3], Catostomidae; [Pattern 4], Cobitidae; [Pattern 5], Cyprinidae; [Pattern 6], Ellopostomatidae; [Pattern 7], Gyriinocheilidae; [Pattern 8], Nemacheilidae; [Pattern 9], Outgroups; [Pattern 10], Vaillantellidae.

the analyses (Table I). Likelihood-ratio tests (Goldman, 1993), implemented in MrModeltest 2.2 (Nylander, 2004), were used to select models for each partition in Partitioned BA (Table I). The parameters for runs in MrBayes included 'lset nst = 6' (GTR) (for partitions 1–9 and 11–18), 'lset nst = 2' (HKY) (for partition 10), 'rates = invgamma' (G+I) (for partitions 1–11, 13 and 16), 'rates = gamma' (G) (for partitions 12, 15, 18), or 'rates = propinv' (I) (for partitions 14 and 17) and '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 Markov-Chain Monte-Carlo (MCMC) chains consisted of 3 000 000 replicates, sampling one tree per 100 replicates. The distribution of log-likelihood scores was examined to determine stationarity for each search and to decide whether extra runs were required to achieve convergence in log-likelihoods among runs or searches. Initial trees with non-stationary log-likelihood values were discarded, and the remaining chains of trees resulting in convergent log-likelihood scores from both independent searches were combined. These trees were used to construct a 50% majority rule consensus tree. For the partitioned ML search with the mixed model of nucleotide substitution, a GTR+G+I model (with four discrete rate categories) for each partition was used as RAxML only provided 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 with 100 separate runs using the default algorithm of the programme from a random starting tree (-d option) for each run. The final tree was selected among suboptimal trees in each run by comparing likelihood scores under the GTR+G+I model.

Nodal support was assessed with bootstrapping (BS) (Felsenstein, 1985) with the maximum likelihood (ML) criterion, based on 1000 pseudo-replicates and the resulting *a posteriori* probabilities from partitioned BA. The MLBS analyses (through analyses using RAxML web servers) (Stamatakis *et al.*, 2008) were conducted with the CIPRES cluster (CIPRES Portal 1-13, http://www.phylo.org/sub_sections/portal/) at the San Diego Supercomputer Center.

RESULTS

A total of 5733 base pairs (bp) were aligned for the exon regions of six nuclear genes in 49 taxa (including four outgroups). Aligned sequence lengths for each locus were 1497 bp (RAG1), 819 bp (RH), 849 bp (IRBP), 846 bp (EGR1), 816 bp (EGR2B) and 906 bp (EGR3). No internal indels appeared in sequences of RAG1, RH and IRBP. A few indels were needed in adjusting sequence alignment of the EGR genes, but the alignment was unambiguously achieved followed by triplet codes for amino acids. Of the 5733 nucleotides, 2821 were variable sites and 2308 of these were parsimony-informative sites. The second or partial RY-coding matrix contained 2313 variable sites, of which 1777 were parsimony-informative sites. Relationships of taxa derived from partitioned ML and Bayesian analyses of DNA sequences based on both matrices were nearly identical with slight differences in relationships within the Cyprinoidea when nodal support was weak; only the ML tree derived from the second (partial RY-coding) matrix is presented (Fig. 2). Most of the resulting clades were highly supported by partitioned MLBS and by *a posteriori* probabilities from partitioned BA (Fig. 2).

In all analyses, the Cypriniformes, Cyprinoidea and all of the cypriniform families represented monophyletic groupings with strong nodal support (Fig. 2). However, these six nuclear genes portrayed the Cobitoidea as a paraphyletic grouping within the order with respect to the Cyprinoidea. Catostomidae formed the basal sister group to the other cypriniform taxa. While this relationship received only weak nodal support in this analysis, the same relationship was also resolved in some other molecular studies that used combined sequence data from three of these nuclear loci,

another nuclear gene encoding growth hormone and whole mt genomes (see Fig. 3 in Mayden *et al.*, 2009).

DISCUSSION

While controversial with respect to the monophyly of the Cobitoidea, the phylogenetic position of the Catostomidae (Nelson, 2006; Saitoh *et al.*, 2006; Šlechtová *et al.*, 2007; Mayden *et al.*, 2009) could not be further resolved here and warrants further evaluation with additional independent molecular markers or morphological data, or both. Within the Cobitoidea (Fig. 2, exclusive of Catostomidae), Gyrinocheilidae formed the sister group to a large, strongly supported clade containing a diverse set of all loach species. The sister-group relationship of the Gyrinocheilidae that was unresolved in previous molecular studies (Saitoh *et al.*, 2006; Mayden *et al.*, 2009) was resolved here, in terms of strong statistical support in trees based on the nuclear DNA sequences in this study (MLBS = 85 and 87%; *a posteriori* probabilities from BA = 1.00 and 1.00 from equal weighting and partial RY-coding analyses, respectively).

The loach clade can be subdivided into six subgroups or lineages corresponding to five currently recognized families plus the newly proposed family Ellopostomatidae (currently containing only species from *Ellopostoma*) recommended herein. Within this large clade, botiid loaches represent a separate lineage from cobitid loaches and are the basal-most members of the loaches (Fig. 2). The non-monophyly for the traditional Cobitidae further corroborates the findings from all available molecular studies (Saitoh *et al.*, 2006; Tang *et al.*, 2006; Šlechtová *et al.*, 2007; Mayden *et al.*, 2008; Mayden *et al.*, 2009). The previously hypothesized shared-derived morphological character of a moveable suborbital spine usually used for aligning the members from these two loach families (Sawada, 1982) will thus likely be found to be homoplastic. The phylogenetic placement of the other enigmatic loach taxa, vailantellid loaches, in relation to other cobitoids was first investigated by Šlechtová *et al.* (2007) using sequence data from a singular nuclear gene, RAG1. As presented in their phylogeny, this family was also resolved herein as the sister group to the other remaining loaches based on variation in these six nuclear genes. This relationship received very high supporting values (MLBS = 100 and 98%; *a posteriori* probabilities from BA = 1.00 and 1.00 from equal weighting and partial RY-coding analyses, respectively) (Fig. 2). Finally, the diverse balitorid and nemacheilid loaches, two families that have long been believed to be closely allied based on the absence of the suborbital spine formed a monophyletic group together with our targeted taxon *Ellopostoma* (Fig. 2). *Ellopostoma* formed the sister group to the nemacheilids, a relationship supported by a moderate MLBS (68%) from equal weighting analysis, but by a higher value (87%) from partial RY-coding analysis (Fig. 2) and by the highest values (1.00) from BA from both types of analyses. This relationship corroborates Kottelat's (1989) earlier morphological hypothesis based on the absence of a suborbital spine and overall similarities, implying that close inspection of morphological variation in these fishes is warranted.

Overall, the results from our present study are consistent with current molecular hypotheses regarding the systematics of Cobitoidea and support the existence of seven previously proposed families plus a new family for a distinct lineage

contains species *Ellopostoma*. Siebert (1987) in his unpublished thesis recognized Ellopostominae as one of the subfamilies in Balitoridae, a hypothesis being rejected herein. Rather, to maintain consistency between phylogenetic relationships of taxa and a natural classification of the organisms, the new family Ellopostomatidae is proposed (following the rules of the International Convention on Zoological Nomenclature, article 13) within the superfamily Cobitoidea. The possible paraphyly of the Cobitoidea remains a vexing problem within the order; despite analysis of complete mitochondrial genomes and now six nuclear genes, evidence for the monophyly or paraphyly of this large clade remains elusive. Further in-depth morphological surveys and analyses may resolve this potential ancient polytomy and/or substantially increased taxon sampling for molecular data may eventually resolve the relationships of these clades (Mayden *et al.*, 2008).

ELLOPOSTOMATIDAE, NEW FAMILY

Diagnosis

The diagnostic characters used for the type genus *Ellopostoma* (Vaillant 1902) (Roberts, 1989; p. 103) are considered valid for diagnosing this family. 'Distinguished from all other cobitids by its oblique, squared-off snout (shaped mainly by enormously expanded maxillae); mouth highly protrusible; a single pair of well-developed maxillary barbels; ceratobranchial 5 with *c.* 30 conical teeth in a single row; suborbital spine absent; dorsal fin elongate, with 18–19 rays, its origin far in advance of a vertical through pelvic-fin origin; pectoral fins not sexually dimorphic; vertebrae 33–34'.

Composition

This family currently includes only two species of *Ellopostoma*: *E. megalomycter* (Vaillant) and *Ellopostoma mystax* Tan & Lim. However, future inventory efforts from Southeast Asia may reveal additional species in this clade and future phylogenetic efforts may identify additional taxa never before examined in a phylogenetic context to be part of this family.

Distribution

Currently, this family is only known from Southeast Asia from peninsular Malaysia, western Borneo and Peninsular Thailand. However, future inventory efforts will likely find additional related taxa in other surrounding and intervening geographic areas.

We thank K. L. Conway and K. L. Tang for providing valuable comments on this study. We thank M. Miya, T. Sado and K. Saitoh for some specimens. We also thank L.H. Chen for the improvement of art illustrations. This research is part of an ongoing international U.S.A. National Science Foundation Tree of Life initiative on the order Cypriniformes to R.L.M. (EF 0431326). Finally, we acknowledge assistant editor S. Grant and two anonymous referees for their constructive comments.

References

- Banareescu, P. M. & Nalbant, T. T. (1995). A generical classification of Nemacheilinae with description of two new genera (Teleostei: Cypriniformes: Cobitidae). *Travaux du Muséum National d'Histoire Naturelle Grigore Antipa* **35**, 429–495.

- 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. *Molecular Phylogenetics and Evolution* **26**, 262–288.
- Chen, W.-J. & Mayden, R. L. (2009). Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the World's largest clade of freshwater fishes: further evidence from six nuclear genes. *Molecular Phylogenetics and Evolution* **52**, 544–549.
- Chen, W.-J., 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.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- Goldman, N. (1993). Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* **36**, 182–198.
- He, S., Gu, X., Mayden, R. L., Chen, W.-J., Conway, K. W. & Chen, Y. (2008). Phylogenetic position of the enigmatic genus *Psilorhynchus* (Ostariophysi: Cypriniformes): evidence from the mitochondrial genome. *Molecular Phylogenetics and Evolution* **47**, 419–425.
- Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES. Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755.
- Kottelat, M. (1989). Zoogeography of the fishes from Indochinese inland waters with an annotated check-list. *Bulletin Zoologisch Museum* **12**, 1–54.
- Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution* **11**, 605–612.
- Mabee, P. M., Arratia, G., Coburn, M., Haendel, M., Hilton, E. J., Lundberg, J. G., Mayden, R. L., Rios, N. & Westerfield, M. (2007). Connecting evolutionary morphology to genomics using ontologies: a case study from Cypriniformes including zebrafish. *Journal of Experimental Zoology B* **308**, 655–668.
- Mayden, R. L., Chen, W.-J., Bart, H. L., Doosey, M. H., Simons, A. M., Tang, K. L., Wood, R. M., Agnew, M. K., Yang, L., Hirt, M. V., Clements, M. D., Saitoh, K., Sado, T., Miya, M. & Nishida, M. (2009). Reconstructing the phylogenetic relationships of the Earth's most diverse clade of freshwater fishes – Order Cypriniformes (Actinopterygii: Ostariophysi): a case study using multiple nuclear loci and the mitochondrial genome. *Molecular Phylogenetics and Evolution* **51**, 500–514.
- Mayden, R. L., Tang, K. L., Conway, K. W., Freyhof, J., Chamberlain, S., Haskins, M., Schneider, L., Sudkamp, M., Wood, R. M., Agnew, M., Bufalino, A., Sulaiman, Z., Miya, M., Saitoh, K. & He, S. (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* **308**, 642–654.
- Mayden, R. L., Tang, K. L., Wood, R. M., Chen, W.-J., Agnew, M. K., Conway, K. W., Yang, L., Simons, A. M., Bart, H. L., Harris, P. M., Li, J., 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. *Journal of Systematics and Evolution* **46**, 424–438.
- Nelson, J. S. (2006). *Fishes of the World*. Hoboken, NJ: John Wiley & Sons, Inc.
- 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. *Molecular Phylogenetics and Evolution* **7**, 247–265.
- Roberts, T. R. (1972). An attempt to determine the systematic position of *Ellopostoma megalomycter*, an enigmatic freshwater fish from Borneo. *Breviaria* **384**, 1–16.
- Roberts, T. R. (1989). *The Freshwater Fishes of Western Borneo (Kalimantan Barat, Indonesia)*. San Francisco, CA: California Academy of Sciences.
- Saitoh, K., Sado, T., Mayden, R. L., Hanzawa, N., Nakamura, K., Nishida, M. & Miya, M. (2006). Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): the first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. *Journal of Molecular Evolution* **63**, 826–841.

- Sawada, Y. (1982). Phylogeny and zoogeography of the superfamily Cobitoidea (Cyprinoidei, Cypriniformes). *Memoirs of the Faculty of Fisheries, Hokkaido University* **28**, 65–223.
- Schilling, T. F. & Webb, J. (2007). Considering the zebrafish in a comparative context. *Journal of Experimental Zoology B* **308**, 515–522.
- Siebert, D. J. (1987). *Interrelationships Among Families of the Order Cypriniformes (Teleostei)*. PhD Thesis, City University of New York, New York, NY, U.S.A.
- Šlechtová, V., Bohlen, J. & Tan, H. H. (2007). Families of Cobitoidea (Teleostei: Cypriniformes) as revealed from nuclear genetic data and the position of the mysterious genera *Barbucca*, *Psilorhynchus*, *Serpenticobitis* and *Vaillantella*. *Molecular Phylogenetics and Evolution* **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. *Systematic Biology* **57**, 758–771.
- Swofford, D. L. (2002). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sunderland, MA: Sinauer Associates.
- Tan, H. H. & Lim, K. K. P. (2002). A new species of *Ellopostoma* (Teleostei: Cypriniformes: Balitoridae) from Peninsular Thailand. *The Raffles Bulletin of Zoology* **50**, 453–457.
- Tang, Q., Liu, H., Mayden, R. L. & 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.
- Vaillant, M. L. (1902). Résultats zoologiques de l'Expédition scientifique néerlandaise au Bornéo central. Note I. *Notes from Leyden Museum* **24**, 1–166.

Electronic References

- Eschmeyer, W. N. & Fricke, R. (2008). *Catalog of Fishes Electronic Version (Updated 18 December 2008)*. San Francisco, CA: California Academy of Sciences. Available at: <http://research.calacademy.org/ichthyology/catalog/fishcatsearch.html>
- Nylander, J. A. A. (2004). MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University. Available at <http://www.abc.se/nnylander/mrmodeltest2.html>

APPENDIX. Taxa included in this study and accession numbers of sequences in GenBank

Family/subfamily	Taxon	GenBank accession number					
		RAG1	RH	IRBP	EGR1	EGR2B	EGR3
Outgroups							
Gonorynchidae	<i>Gonorynchus greyi</i>	EU409606	EU409632	EU409664	EU409696	EU409728	EU409760
Bagridae	<i>Pseudobagrus tokiensis</i>	FJ650410	FJ197075		FJ650422	FJ650438	FJ650454
Characidae	<i>Chalceus macrolepidotus</i>	EU409607	EU409633	EU409665	EU409697	EU409729	EU409761
Alestiidae	<i>Phenacogrammus interruptus</i>	FJ197124	FJ197073	FJ197123	FJ650423	FJ650439	FJ650455
Cypriniformes							
Cobitoidea							
Balitoridae	<i>Sewellia lineolata</i>	EU409609	EU409635	EU409667	EU409699	EU409731	EU409763
Balitoridae	<i>Homaloptera parvitella</i>	EU409610	EU409636	EU409668	EU409700	EU409732	EU409764
Catostomidae	<i>Catostomus commersoni</i>	EU409612	EU409638	EU409670	EU409702	EU409734	EU409766
Catostomidae	<i>Cycleptus elongatus</i>	EU409613	EU409639	EU409671	EU409703	EU409735	EU409767
Botiidae	<i>Botia dario</i>	EU409614	EU409641	EU409673	EU409705	EU409737	EU409769
Botiidae	<i>Leptobotia pellegrini</i>	EU292683	EU409640	EU409672	EU409704	EU409736	EU409768
Botiidae	<i>Syncrossus beauforti</i>	FJ650411	FJ650470	FJ650482	FJ650424	FJ650440	FJ650456
Botiidae	<i>Yasuhikotakia morleti</i>	FJ650412	FJ650471	FJ650483	FJ650425	FJ650441	FJ650457
Cobitidae	<i>Acaniopsis</i> sp.	FJ650413	FJ650472	FJ650484	FJ650426	FJ650442	FJ650458
Cobitidae	<i>Cobitis takatsiensis</i>	EU409616	EU409643	EU409675	EU409707	EU409739	EU409771
Cobitidae	<i>Niwaella multifasciata</i>	EU409615	EU409642	EU409674	EU409706	EU409738	EU409770
Cobitidae	<i>Pangio oblonga</i>	EU711141	FJ197041	FJ197091	FJ650427	FJ650443	FJ650459
Cobitidae	<i>Somileptus gongota</i>	FJ650414	FJ650473	FJ650485	FJ650428	FJ650444	FJ650460
Gyrinocheilidae	<i>Gyrinocheilus aymonieri</i>	EU292682	FJ197071	FJ197122	EU409727	EU409759	EU409791
Gyrinocheilidae	<i>Gyrinocheilus pennocki</i>	FJ650415	FJ650474	FJ650486	FJ650429	FJ650445	FJ650461
Nemacheilidae	<i>Acanthocobitis</i> sp.	FJ650416	FJ650475	FJ650487	FJ650430	FJ650446	FJ650462
Nemacheilidae	<i>Barbatula barbata</i>	EU711107	FJ650476	FJ650488	FJ650431	FJ650447	FJ650463
Nemacheilidae	<i>Lefua costata</i>	EU409608	EU409634	EU409666	EU409698	EU409730	EU409762
Nemacheilidae	<i>Oreonectes platycephalus</i>	FJ650418	FJ650478	FJ650490	FJ650433	FJ650449	FJ650465

APPENDIX. Continued

Family/subfamily	Taxon	GenBank accession number						
		RAG1	RH	IRBP	EGR1	EGR2B	EGR3	
Nemacheilidae	<i>Schistura savona</i>	FJ650419	FJ650479	FJ650491	FJ650434	FJ650450	FJ650466	
Nemacheilidae	<i>Traccatichthys pulcher</i>	EU409611	EU409637	EU409669	EU409701	EU409733	EU409765	
Nemacheilidae	<i>Triplophysa gundriseri</i>	FJ650420	FJ650480	FJ650492	FJ650435	FJ650451	FJ650467	
Nemacheilidae	<i>Tuberoschistura baenzigeri</i>	FJ650421	FJ650481	FJ650493	FJ650436	FJ650452	FJ650468	
Ellopostomatidae	<i>Ellopostoma mystax</i>	FJ650417	FJ650477	FJ650489	FJ650432	FJ650448	FJ650464	
Vaillantellidae	<i>Vaillantella maassi</i>	EU711132	FJ197031	FJ197080	FJ650437	FJ650453	FJ650469	
Cyprinidae	<i>Acheilognathus tabira</i>	EU409617	EU409644	EU409676	EU409708	EU409740	EU409772	
Cyprinidae	<i>Tanaka himantegus</i>	EU409618	EU409645	EU409677	EU409709	EU409741	EU409773	
Cyprinidae	<i>Ischikauia steenackeri</i>	EU292687	EU409648	EU409680	EU409712	EU409744	EU409776	
Cyprinidae	<i>Megalobrama amblycephala</i>	EU409620	EU409647	EU409679	EU409711	EU409743	EU409775	
Cyprinidae	<i>Garra spilota</i>	EU409621	EU409649	EU409681	EU409713	EU409745	EU409777	
Cyprinidae	<i>Hampala macrolepidota</i>	EU409623	EU409651	EU409683	EU409715	EU409747	EU409779	
Cyprinidae	<i>Labeo chrysopehkadion</i>	EU409622	EU409650	EU409682	EU409714	EU409746	EU409778	
Cyprinidae	<i>Bivia zezera</i>	EU409626	EU409654	EU409686	EU409718	EU409750	EU409782	
Cyprinidae	<i>Romanogobio ciscaucasicus</i>	EU409624	EU409652	EU409684	EU409716	EU409748	EU409780	
Cyprinidae	<i>Sarcocheilichthys parvus</i>	EU409625	EU409653	EU409685	EU409717	EU409749	EU409781	
Cyprinidae	<i>Notropis baileyi</i>	EU292691	EU409657	EU409689	EU409721	EU409753	EU409785	
Cyprinidae	<i>Phoxinus perenurus sachalinensis</i>	EU409627	EU409655	EU409687	EU409719	EU409751	EU409783	
Cyprinidae	<i>Scardinus erythrophthalmus</i>	EU409628	EU409656	EU409688	EU409720	EU409752	EU409784	
Cyprinidae	<i>Semotilus atromaculatus</i>	EU409629	EU409658	EU409690	EU409722	EU409754	EU409786	
Cyprinidae	<i>Danio albolineatus</i>	EU292696	EU409661	EU409693	EU409725	EU409757	EU409789	
Cyprinidae	<i>Danio dangila</i>	EU292697	EU409660	EU409692	EU409724	EU409756	EU409788	
Cyprinidae	<i>Danio rerio</i>	U71093	L11014	X85957	NM_131248	NM_130997	scaffold2320.1	
Cyprinidae	<i>Macrochirichthys macrochirus</i>	EU409630	EU409659	EU409691	EU409723	EU409755	EU409787	
Cyprinidae	<i>Rasbora steineri</i>	EU409631	EU409662	EU409694	EU409726	EU409758	EU409790	
Cyprinidae	<i>Paralabauca typus</i>	EU409619	EU409646	EU409678	EU409710	EU409742	EU409774	