J Mol Evol (2006) 63:826–841
DOI: 10.1007/s00239-005-0293-y

JOURNAL OF MOLECULAR EVOLUTION

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

K. Saitoh,[1] T. Sado,[2] R. L. Mayden,[3] N. Hanzawa,[4] K. Nakamura,[4] M. Nishida,[5] M. Miya[2]

[1] Tohoku National Fisheries Research Institute, Miyagi 985-0001, Japan
[2] Natural History Museum and Institute, Chiba, Chiba 260-8682, Japan
[3] Department of Biology, Saint Louis University, St. Louis, MO 63103-2010, United States
[4] Department of Biology, Yamagata University, Yamagata 990-8560, Japan
[5] Ocean Research Institute, University of Tokyo, Tokyo 164-8639, Japan

Received: 3 December 2005 / Accepted: 21 July 2006 [*Reviewing Editor:* Rafael Zardoya]

Abstract. Fishes of the order Cypriniformes are almost completely restricted to freshwater bodies and number > 3400 species placed in 5 families, each with poorly defined subfamilies and/or tribes. The present study represents the first attempt toward resolution of the higher-level relationships of the world's largest freshwater-fish clade based on whole mitochondrial (mt) genome sequences from 53 cypriniforms (including 46 newly determined sequences) plus 6 outgroups. Unambiguously aligned, concatenated mt genome sequences (14,563 bp) were divided into 5 partitions (first, second, and third codon positions of the protein-coding genes, rRNA genes, and tRNA genes), and partitioned Bayesian analyses were conducted, with protein-coding genes being treated in 3 different manners (all positions included; third codon positions converted into purine [R] and pyrimidine [Y] [RY-coding]; third codon positions excluded). The resultant phylogenies strongly supported monophyly of the Cypriniformes as well as that of the families Cyprinidae, Catostomidae, and a clade comprising Balitoridae + Cobitidae, with the 2 latter loach families being reciprocally paraphyletic. Although all of the data sets yielded nearly identical tree topologies with regard to the shallower relationships, deeper relationships among the 4 major clades (the above 3 major clades plus Gyrinocheilidae, represented by a single species *Gyrinocheilus aymonieri* in this study), were incongruent depending on the data sets. Treatment of the rapidly saturated third codon–position transitions appeared to be a source of such incongruities, and we advocate that RY-coding, which takes only transversions into account, effectively removes this likely "noise" from the data set and avoids the apparent lack of signal by retaining all available positions in the data set.

Key words: Evolutionary genomics — Partitioned Bayesian analysis — RY-coding — Third codon position — Treeness/RCV

Introduction

Resolving higher-level relationships of highly speciose lineages remains a major challenge in phylogeny reconstruction because it requires a tremendous amount of both taxon and character sampling. In molecular systematics of fishes, technical difficulties in character sampling have been partly alleviated with the advent of a polymerase chain reaction (PCR)–

Correspondence to: M. Miya; *email:* miya@chiba-muse.or.jp

based approach for sequencing whole mitochondrial (mt) genomes (Miya and Nishida 1999), producing strictly homologous characters that are prerequisite for correct estimation of phylogenies (Martin & Burg 2002). With purposeful, extensive taxon sampling, whole mt genome sequences have resolved many fish phylogenies at various taxonomic levels with strong statistical support (Inoue et al. 2001c, 2003, 2004; Miya et al. 2001, 2003, 2005; Ishiguro et al. 2003, 2005; Saitoh et al. 2003; Yamanoue et al. 2004; Lavoué et al. 2005; Minegishi et al. 2005), and the resulting trees are highly congruent with those derived from independent nuclear markers (Lê et al. 1993; Wiley et al. 2000; Zaragueta-Bagils et al. 2002; Chen et al. 2003; López et al. 2004; Smith and Wheeler 2004; Dettai and Lecointre 2005). Despite these efforts, many phylogenetic questions remain unanswered for many diverse clades of fishes wherein diversity exceeds several hundreds to thousands of species.

The order Cypriniformes (minnows and their relatives) comprising >3400 species (FishBase July 2005; http://www.fishbase.org; Table 1) represents an excellent example of a highly diverse clade with tremendous species diversity. With the exception of undoubtedly polyphyletic Perciformes (Johnson and Patterson 1993; Miya et al. 2003, 2005; Chen et al. 2003; Dettai and Lecointre 2005), Cypriniformes is the largest order of all fishes, followed by Siluriformes (catfishes), with comparable species' diversity (approximately 2930 spp.; FishBase July 2005). Unlike the Siluriformes, however, which is divided into 34 families and 412 genera (Nelson 1994), the Cypriniformes is divided into only 5 families (Cyprinidae, minnows; Catostomidae, suckers; Gyrinocheilidae, algae eaters; Cobitidae, loaches Balitoridae, river loaches), each with poorly defined subfamilies (e.g., 2 to 12 subfamilies recognized in cyprinids depending on investigators; Arai 1982; Chen et al. 1984; Howes 1991; Cavender and Coburn 1992) and 426 genera (Table 1). This skewed classification is partly related to taxonomic analyses relying either on external (e.g., presence or absence of barbel) or internal (e.g., pharyngeal teeth) characters (Howes 1991) and reflects clear difficulties in recognizing monophyletic groups solely based on morphological attributes. Furthermore, such enormous species diversity has precluded comprehensive phylogenetic analyses based on both morphological and molecular characters. Accordingly, most studies of large taxonomic scope have necessarily included restricted numbers of terminal taxa (Wu et al. 1981; Siebert 1987) or are restricted to individual families from limited areas (Sawada 1982; Chen et al. 1984; Howes 1991; Cavender and Coburn 1992; Coburn and Cavender 1992; Smith 1992; Briolay et al. 1998; Gilles et al. 1998, 2001; Zardoya and Doadrio 1998, 1999a, b; Hänfling and Brandl 2000; Harris and Mayden 2001; Okazaki et al. 2001; Xiao et al. 2001; Cunha et al. 2002; Durand et al. 2002; Liu and Chen 2003; Simons et al. 2003; Clements et al. 2004; He et al. 2004a; Tang et al. 2005). In particular, although our previous mt genomic study corroborated the placement of some cypriniform taxa among ostariophysan orders (Saitoh et al. 2003), it should be noted that no molecular phylogenetic study has been published with extensive taxon and character sampling from all representative lineages from the cypriniforms.

Cypriniform fishes are found in a wide spectrum of freshwater environments of all continents except for Oceania, Antarctic, and South America (Berra 2001). They occur in cold streams near the Arctic, rivers in

Table 1. Diversity of the Cypriniformes expressed by number of genera and species in each family and/or subfamily and number of species used in this study

Family	Subfamily[a]	No. genera[b]	No. species[b]	This study
Cyprinidae	Cyprininae	10	104	2
	Barbinae	46	756	4
	Labeoninae	27	355	2
	Schizothoracinae	17	103	1
	Acheilognathinae	6	61	2
	Gobioninae	30	179	6
	Leuciscinae[c]	38	204	3
	Phoxininae	48	319	4
	Xenocyprininae	9	16	1
	Cultrinae	15	63	2
	Rasborinae	51	318	5
	Tincinae	1	1	1
	Psilorhyncinae	2	6	–
	Incertae sedis	24	37	–
	Subtotal	324	2522	33
Gyrinocheilidae		1	3	1
Catostomidae	Catostominae	10	65	4
	Cycleptinae	2	3	2
	Ictiobinae	2	8	1
	Subtotal	14	76	7
Cobitidae	Cobitinae	20	146	4
	Botinae	3	55	2
	Subtotal	23	201	6
Balitoridae	Balitorinae	28	168	2
	Nemacheilinae	31	443	4
	Incertae sedis	5	6	–
	Subtotal	64	617	6
Total		426	3419	53

[a] Cyprinid subfamilial classification largely corresponds to informal taxa recognized by Cavender and Coburn (1992) within their 2 subfamilies Cyprininae and Leuciscinae. Schizothoracinae (= schizothoracin lineage) was added according to Howes (1991). For genera unassigned to any of the subfamilies in Cavender and Coburn (1992), Chen et al. (1994), Talwar and Jhingran (1991), and Cunha et al. (2002) were referred. Nevertheless, 37 species belonging to 24 genera were unassignable. Catostomid subfamilial classification follows Smith (1992). Cobitid and balitorid subfamilial classifications follows Sawada (1982), and assignment of genera follows Nelson (1994).

[b] Data were taken from FishBase (http://www.fishbase.org; July 2005).

[c] Including Alburninae sensu Howes (1991).

tropical rainforests, Himalayan highland-glacial lakes and torrents, lowland floodplain swamps, lakes, rivers, artificial reservoir, creeks, and rice paddies, etc. (FishBase; http://www.fishbase.org). Their sizes vary greatly from one of the smallest vertebrates (*Paedocypris progenetica*, 7.9 mm in standard length [SL] [Kottelat et al. 2005]) to one of the biggest freshwater fishes (*Catlocarpio siamensis*, 3.0 m SL or 300 kg [Baird et al. 1999; Roberts and Warren 1994]) in the world. Cypriniform fishes are also known as a fish group with frequent occurrence of both bisexual and unisexual polyploid populations (*Cyprinus*, *Carassius*, *Barbus*, *Cobitis*, *Botia*, entire Catostomidae, etc. [Vrijenhoek et al. 1989; Leggatt and Iwama 2003]), some of which even display cyclic ploidy shifts between generations (*Squalius alburnoides* and *Cobitis hankugensis-longicorpus* complexes [Alves et al. 2002; Saitoh et al. 2004]). Cypriniform fishes also include zebrafish (*Danio rerio*), a representative model organism. This species and its mutants have been extensively used in developmental and genetic research (Kimmel 1989; Duyk and Schmitt 2001; see also ZFIN at http://www.zfin.org). The fathead minnow (*Pimephales promelas*) is widely used for environmental testing of various chemical substances (PAN Pesticides Database at http://www.pesticideinfo.org) and in experimental animal of behavioral ecology (e.g., Mirza and Chivers 2003).

In this article, we address the following 3 phylogenetic questions regarding the higher-level relationships of the order Cypriniformes: (1) monophyly of the Cypriniformes; (2) monophyly of the 4 families (Cyprinidae, Catostomidae, Cobitidae, and Balitoridae); and (3) interrelationships among the resultant major lineages (at a level of family or combinations of families). In addition, we examine newly discovered monophyletic groups within these major lineages with respect to the previously recognized subfamilies or tribes.

Materials and Methods

Taxon Sampling

For resolution of a complex phylogeny with enormous taxonomic diversity, such as that seen in the Cypriniformes (Table 1), it is essential to conduct purposeful taxon sampling that increases phylogenetic accuracy (Hillis 1998). We employed 2 taxon-sampling strategies, individually or in combination, according to the following recommendations: (1) "select taxa within the monophyletic group of interest that will represent the overall diversity of the group" (strategy no. 3 in Hillis [1998]); and (2) "select taxa within the monophyletic group of interest that are expected (based on current taxonomy or previous phylogenetic studies) to subdivide long branches in the initial tree" (strategy no. 4 in Hillis [1998]). Hillis (1998) stated that careful addition of taxa to ensure coverage of the group of interest and to purposefully break up long branches (a combination of strategy nos. 3 and 4) seemed to be the optimal taxon-sampling strategy.

For the former recommendation (strategy no. 3), we chose at least 1 species from each family or subfamily (Table 1) except for the Psilorhynchinae, for which a tissue sample was unavailable. For the latter recommendation (strategy no. 4), species were chosen that would bisect the resulting long branches observed in Saitoh et al. (2003) and that subsequently were found in a series of preliminary phylogenetic analyses concurrently conducted with sequencing experiments. Consequently, in addition to the previously determined 7 species of cypriniforms (Saitoh et al. 2003), we determined the whole mt genome sequences for an additional 46 species. All species used in this study are listed in Table 2 along with references and DDBJ/EMBL/GenBank accession numbers.

DNA Extraction, PCR, and Sequencing

Extraction of total genomic DNA followed the standard phenol-chloroform method as described in Asahida et al. (1996). The mt genomes of the 46 cypriniform species were amplified in their entirety using a long PCR technique (Cheng et al. 1994). We used 6 fish-versatile long PCR primers in the following 4 combinations (S-LA-16S-L + S-LA-16S-H; L2508-16S + H1065-12S; L2508-16S + H12293-Leu; and L12321-Leu + S-LA-16S-H; for locations and sequences of these primers, see Miya and Nishida 1999, 2000; Inoue et al. 2000, 2001c; Ishiguro et al. 2001; Kawaguchi et al. 2001) so as to amplify the entire mt genome in a single or 2 reactions. Long PCR reaction conditions followed Miya and Nishida (1999). Dilution of the long PCR products with sterile water (1:10 to 100) subsequently served as templates for short PCR reactions.

We used a total of 162 fish-versatile PCR primers in various combinations to amplify contiguous, overlapping segments of the entire mt genome for each of the 46 species (for locations and sequences of the primers, see Miya and Nishida 1999, 2000; Inoue et al. 2000, 2001a, 2001b, 2001c, 2001d; Ishiguro et al. 2001; Kawaguchi et al. 2001). In addition, we used 14 (Miya et al. 2006) and 84 newly-designed cypriniform-specific primers for short PCRs and sequencing to accommodate the cypriniform mt genomes in later stages of this study. Short PCR reaction conditions followed Miya and Nishida (1999). A list of PCR primers used for a specific species is available from K. S. on request.

Double-stranded short PCR products, purified using a PreSequencing kit (USB), subsequently served for direct cycle sequencing with dye-labeled terminators (Applied Biosystems and Amersham Pharmacia) with the same primers for the short PCRs. All sequencing reactions were performed according to the manufacture's instructions. Labeled fragments were analyzed on model 373/377/3100 sequencers (Applied Biosystems).

Alignment

The character matrix constructed in the previous study (Saitoh et al. 2003) formed the basis of the present alignments for 13 protein-coding and 22 tRNA genes. We manually aligned newly determined sequences with reference to deduced amino-acid sequences for the 13 protein-coding genes and secondary structure models for 2 rRNA (De Rijk et al. 1998; Van de Peer et al. 1998) and 22 tRNA (Kumazawa and Nishida 1993) genes. We excluded all gaps and ambiguous alignment sites, resulting in 11,175, 2,070, and 1,318 nucleotide positions (total = 14,563 positions) from the 13 protein-coding, 2 rRNA, and 22 tRNA genes, respectively. All sequences from the L-strand–encoded genes (ND6 and 8 tRNA genes) were converted into complementary strand sequences.

We dealt with the protein-coding gene sequences as follows to construct 3 different data sets (all including rRNA and tRNA gene sequences): (1) all positions included (designated as 123_nRT_n; subscript "n" denotes nucleotides); (2) third codon positions con-

Table 2. List of species used in this study with DDBJ/EMBL/GenBank accession numbers and references

Classification	Species	Accession No.	References
Order Gonorynchiformes			
Family Chanidae	*Chanos chanos*	AB054133	Saitoh et al. (2003)
Family Gonorynchidae	*Gonorynchus greyi*	AB054134	Saitoh et al. (2003)
Order Siluriformes			
Family Bagridae	*Pseudobagrus tokiensis*	AB054127	Saitoh et al. (2003)
Family Callichthyidae	*Corydoras rabauti*	AB054128	Saitoh et al. (2003)
Order Characiformes			
Family Alestiidae	*Phenacogrammus interruptus*	AB054129	Saitoh et al. (2003)
Family Characidae	*Chalceus macrolepidotus*	AB054130	Saitoh et al. (2003)
Order Cypriniformes			
Family Cyprinidae			
Subfamily Cyprininae	*Carassius auratus langsdorfii*	AB006953	Murakami et al. (1997)
	Cyprinus carpio	AP009047	Mabuchi et al. (2006)
Subfamily Barbinae	*Barbus barbus*	AB238965	This study
	Barbus trimaculatus	AB239600	This study
	Barbonymus gonionotus	AB238966	This study
	Puntius ticto	AB238969	This study
Subfamily Labeoninae	*Labeo batesii*	AB238967	This study
	Labeo senegalensis	AB238968	This study
Subfamily Schizothoracinae	*Gymnocypris przewalskii*	AB239595	This study
Subfamily Acheilognathinae	*Acheilognathus typus*	AB239602	This study
	Rhodeus ocellatus kurumeus	AB070205	This study
Subfamily Gobioninae	*Gobio gobio*	AB239596	This study
	Gnathopogon elongates	AB218687	This study
	Hemibarbus barbus	AB070241	This study
	Sarcocheilichthys variegatus mircrooculus	AB054124	Saitoh et al. (2003)
	Pseudorasbora pumila pumila	AB239599	This study
	Pungtungio herzi	AB239598	This study
Subfamily Leuciscinae	*Alburnus alburnus*	AB239593	This study
	Notemigonus crysoleucas	AB127393	This study
	Pelecus clutratus	AB239597	This study
Subfamily Phoxininae	*Cyprinella lutrensis*	AB070206	This study
	Phoxinus perenurus	AP009061	This study
	Pseudaspius lectocephalus	AP009058	This study
	Tribolodon nakamurai	AB218896	This study
Subfamily Xenocyprinae	*Xenocypris argentea*	AP009059	This study
Subfamily Cultrinae	*Culter mongolicus*	AP009060	This study
	Ischikauia steenackeri	AB239601	This study
Subfamily Rasborinae	*Aphyocypris chinensis*	AB218688	This study
	Danio rerio	AC024175	Broughton et al. (2001)
	Esomus metallicus	AB239594	This study
	Opsariichthys uncirostris	AB218897	This study
	Zacco sieboldii	AB218898	This study
Incertae sedis	*Tinca tinca*	AB218686	This study
Family Catostomidae			
Subfamily Catostominae	*Catostomus commersonii*	AB127394	This study
	Minytrema melanops	AB242166	This study
	Moxostoma poecilurum	AB242167	This study
	Hypentelium nigricans	AB242169	This study
Subfamily Cycleptinae	*Myxocyprinus asiaticus*	AP006764	This study
	Cycloptus elongates	AB126082	This study
Subfamily Ictiobinae	*Carpiodes carpio*	AP006763	This study
Family Gyrinocheilidae	*Gyrinocheilus aymonieri*	AB242164	This study
Family Cobitidae			
Subfamily Cobitinae	*Cobitis striata*	AB054125	Saitoh et al. (2003)
	Misgurnus nikolskyi	AB242171	This study
	Pangio anguillaris	AB242168	This study
	Acantopsis choirorhynchos	AB242161	This study
Subfamily Botinae	*Botia macracanthus*	AB242163	This study
	Leptobotia mantschurica	AB242170	This study
Family Balitoridae			
Subfamily Baritorinae	*Homaloptera leonardi*	AB242165	This study
	Crossostoma lacustre	M91245	Tzeng et al. (1992)

Continued

Table 2. Continued

Classification	Species	Accession No.	References
Subfamily Nemacheilinae	*Barbatula toni*	AB242162	This study
	Lefua echigonia	AB054126	Saitoh et al. (2003)
	Schistura balteata	AB242172	This study
	Vaillantella maassi	AB242173	This study

verted into purine (R) and pyrimidine (Y) (RY-coding; designated as $12_n3_rRT_n$; subscript "r" denotes RY-coding; Phillips and Penny 2003; Harrison et al. 2004); and (3) third codon positions excluded (designated as 12_nRT_n). Aligned sequence data in NEXUS format are available from K. S. on request.

Analysis of Sequence Variations

Pairwise comparisons and statistical information from the mt genome sequences were obtained using PAUP 4.0b10 (Swofford 2001). To examine patterns of sequence variations for the first, second, and third codon positions of the protein-coding genes, rRNA, and tRNA genes separately, we plotted pairwise nucleotide differences (sorted into transitional [TS] and transversional [TV] differences) against an evolutionary distance as a substitute for absolute geological time. The gamma-corrected maximum-likelihood (ML) distance with the mtREV model (Adachi and Hasegawa 1996), derived from deduced amino-acid sequences from the 13 protein-coding genes, which has been demonstrably linear with the absolute geological time for selected vertebrate taxa (Kumazawa et al. 2004), was calculated with TREE-PUZZLE 5.2 (Schmidt et al. 2002) and used as the evolutionary distance.

Phylogenetic Analysis

When using an mt genome data set with a purposeful taxon sampling, Simmons and Miya (2004) empirically showed that Bayesian analysis (Rannala and Yang 1996; Yang and Rannala 1997; Larget and Simon 1999; Mau et al. 1999) is the most efficient character-based method for accurately reconstructing phylogeny. According to their recommendations, we used this method for the 3 different data sets.

Partitioned Bayesian phylogenetic analyses were conducted with a parallel version of MrBayes 3.0b4 (Ronquist and Huelsenbeck 2003). The program was run using 2 separate clusters of 4 Linux nodes, each equipped with Pentium 4 (3.0 GHz). We set 5 (123_nRT_n and $12_n3_rRT_n$) and 4 (12_nRT_n) partitions depending on the data sets, assuming that functional constraints on sequence evolution are more similar within codon positions (or types of molecules) across genes than across codon positions (or types of molecules) within genes, at least for a set of mt genes.

The general time-reversible model with some sites assumed to be invariable and with variable sites assumed to follow a discrete gamma distribution (GTR + I + Γ; Yang 1994) was consistently selected as the best-fit model of nucleotide substitution by MrModeltest2 (http://www.ebc.uu.se/systzoo/staff/nylander) for each partition except for positions with the RY-coding. We assumed that all of the model parameters were unlinked and that the rate multipliers were variable across partitions. We used the default settings for the priors on the proportion of invariable site (0 to 1) and the gamma shape parameter (0.1 to 50.0). A Dirichlet distribution was assumed for the rate matrix and base frequency, and every tree topology was assumed to be equally probable.

For third codon positions in $12_n3_rRT_n$ coding, we used arbitrarily "A" and "C" instead of "R" and "Y" and set a single rate category (lset nst = 1) instead of 6 (lset nst = 6) to allow the program to estimate only transversional changes between purine (R) and pyrimidine (Y) nucleotides. RY-coding originally intends to use only transversional changes, although MrBayes unnecessarily estimated transitional changes (A↔G or C↔T) when RY-coding was employed, which simply imposed fixed nucleotide compositions on R (A + G) and Y (C + T) during the calculation.

The Markov chain Monte Carlo (MCMC) process was set so that 4 chains (3 heated and 1 cold) ran simultaneously. We conducted 4 independent runs for each data set and continued the runs for 3.0×10^6 cycles, with 1 in every 1000 trees being sampled. In all runs, "stationarity" (lack of improvement in the likelihood scores) was reached no later than 3.0×10^5 cycles, and we discarded the first 301 trees as a "burn-in" period. All parameters estimated from the remaining 2,700 trees from the 4 independent runs were in excellent agreement for each of the 3 data sets. Thus, posterior probabilities (PPs) of the phylogenies, and their branches were determined based on the 10,800 trees pooled from the 4 runs for the 3 data sets.

Signal-to-Noise Ratio

To check the phylogenetic signal in different data sets and partitions, we calculated "relative compositional variability" (RCV; for details, see later) values on the data and treeness values on the tree with the highest probability in the $12_n3_rRT_n$ data set. We used this tree topology as a starting tree and constrained a tree topology using a "prset topologypr" command. By specifying a starting tree and any arbitrary constraint, MrBayes 3.04 searches parameters that maximize the posterior probability on that single tree topology. We conducted a single run for each data set and partition according to the method described previously except for the number of runs [1 instead of 4], cycles [1.0 to 5.0×10^5 instead of 3.0×10^6], and a burn-in period [31 instead of 301 trees]). Treeness statistics is the sum of internal branch lengths divided by the sum of all branch lengths (Lanyon 1988; Phillips and Penny 2003; Harrison et al. 2004). Treeness increases when apparently saturated sites are omitted, such as third codon positions, or when the nucleotides are reduced to pyrimidines and purines (Harrison et al. 2004). Such a reduced bias is measured by the RCV, i.e., the average variability for character states between taxa. For nucleotides, RCV is defined as:

$$\mathrm{RCV} = \left(\sum |\mathrm{A}_i - \mathrm{A}*| + |\mathrm{C}_i - \mathrm{C}*| + |\mathrm{G}_i - \mathrm{G}*| + |\mathrm{T}_i - \mathrm{T}*|\right)/\mathrm{n} \times \mathrm{t},$$

where A_i, C_i, G_i, and T_i are the total numbers of each nucleotide for the ith taxon; A*, C*, G*, and T* are the averages for the n taxa; and t is the number of sites (Phillips and Penny 2003; Harrison et al. 2004). RCV allows direct comparisons of the extent of composition bias for data sets and data treatments.

Testing Alternative Topologies

An important use of Bayesian analysis is in the consideration of alternative topologies using the collection of trees (Miller et al. 2002). By counting the number of times a particular hypothesis appeared in the simulation, we quantitatively assessed the strength of support for the hypothesis.

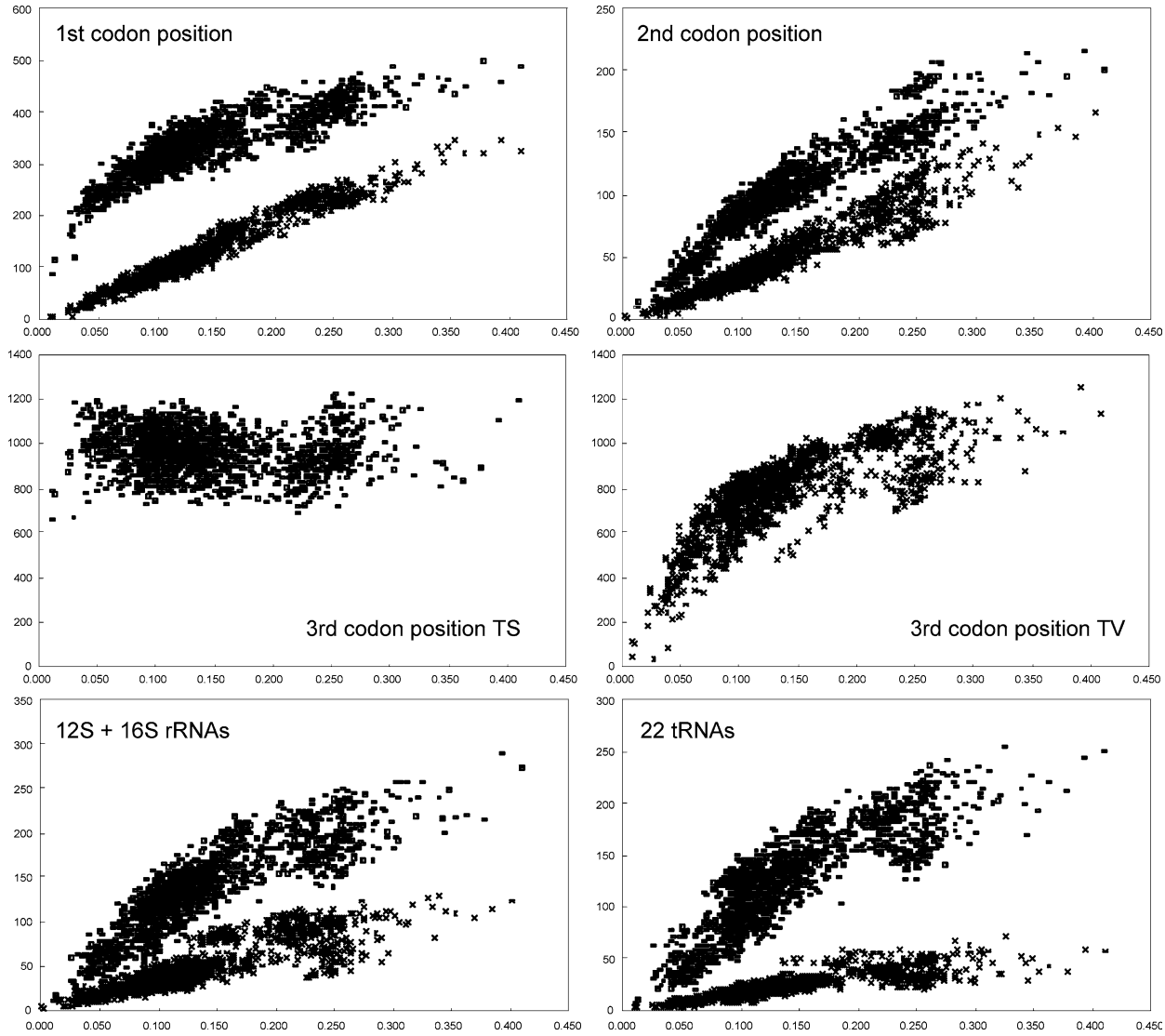

Fig. 1. Patterns of sequence variation for mt genome data from the 53 cypriniforms and 6 outgroups used in this study. Pairwise transitional (closed circle = TS) and transversional (x = TV) differences are plotted against an evolutionary distance as a substitute for absolute geological time. The gamma-corrected maximum likelihood distance with mtREV model (Adachi and Hasegawa 1996), derived from deduced amino-acid sequences from the 13 protein-coding genes, was used for the evolutionary distance.

When alternative hypotheses could be specified in single topologies, we also compared those topologies using the likelihood-based KH (Kishino and Hasegawa 1989) and SH (Shimodaira and Hasegawa 1999) tests as implemented in PAUP. Because PAUP does not allow for data partitioning, we had to deal with each of the 3 concatenated sequences as a single partition. The GTR + I + Γ was selected by MrModeltest2 as the best-fit model of sequence evolution for all 3 data sets, and the estimated parameters were used in all the comparisons. We estimated the variance in likelihood difference between 2 topologies using the resampling estimated log-likelihood method from 10,000 bootstrap replications, and the difference was statistically evaluated.

Results and Discussion

Genome Organization

The complete L-strand nucleotide sequences from the mt genomes of the 46 species reported here for the first time have been registered in DDBJ/EMBL/GenBank (Table 2). The genome content of the 46 species included 13 protein-coding, 2 rRNA, and 22 tRNA genes, plus the putative control region, as found in other vertebrates. As in other vertebrates, most genes were encoded on the H-strand, except for the ND6 and 8 tRNA genes, which were encoded on the L-strand. No unique gene order was found in mt genomes from the 46 species.

Patterns of Sequence Variations

Both the pairwise TS and TV differences for each partition increased with an increase in evolutionary distance, with the exception of TS differences at third codon positions of the protein-coding genes (Fig. 1),

Table 3. Treeness and RCV values for datasets and partitions for the 53 cypriniform taxa used in this study

Dataset/partition[a]	Treeness	RCV	Treeness/RCV
123_nRT_n	0.2095	0.0421	4.976
$12_n3_rRT_n$	0.1892	0.0249	7.598
12_nRT_n	0.2103	0.0158	13.310
1_n	0.1758	0.0097	18.124
2_n	0.2190	0.0058	37.759
3_n	0.2117	0.1172	1.806
3_r	0.1592	0.0751	2.120
R_n	0.2088	0.0153	13.647
T_n	0.2718	0.0162	16.778

[a] 123_nRT_n, all positions included (subscript "n" denotes nucleotides); $12_n3_rRT_n$, third codon positions converted into purine (R) and pyrimidine (Y) (RY-coding; subscript "r" denotes RY-coding; Phillips and Penny 2003; Harrison et al. 2004); 12_nRT_n, 3rd codon positions excluded.

in which marked saturation has been observed in earlier stages of evolution (< 5% evolutionary distance) with no increases being observed thereafter. It was apparent that some degree of saturation had also occurred in other positions (particularly in TS), although the pairwise differences seemed to accumulate steadily along the time axis.

Signal-to-Noise Ratio

For the 3 data sets, there were only slight differences among treeness statistics (0.1892 to 0.2103; Table 3), although RCV decreased from 0.0421 to 0.0158 as third codon positions were reduced to RY ($12_n3_rRT_n$) and third codon positions (12_nRT_n) were omitted entirely from the data set, both resulting in significant increases in the signal-to-noise ratio (treeness/RCV) from 4.976 (123_nRT_n) through 7.598 ($12_n3_rRT_n$) to 13.310 (12_nRT_n) (Table 3).

There were also relatively small differences among treeness statistics from the 6 partitions (0.1592 to 0.2718), although RCV varied greatly from 0.0058 (2_n) to 0.1172 (3_n) (Table 3). Consequently, the second codon position (2_n) exhibited the highest signal-to-noise ratio (treeness/RCV = 37.759), followed by first codon positions (1_n), tRNA (T_n) and rRNA (R_n) (18.124, 16.778, and 13.647, respectively). RY-coding improved the signal-to-noise ratio at third codon positions from 1.806 (3_n) to 2.120 (3_r) (Table 3).

Noteworthy is that the ratio treeness/RCV does not reflect quantity of historical signal in the data sets. Considering the additive nature of historical signals (i.e., historical signals collectively support one correct tree), the use of all available positions in an analysis to avoid a lack of historical signals, although removing expected or likely noise, would be an ideal approach for correct estimation of phylogenies (see later).

Major Patterns of Cypriniform Phylogeny

A 50% majority consensus tree of the 10,800 pooled trees from the 4 independent Bayesian analyses of the $12_n3_rRT_n$ (RY-coding) data set resulted in a fully bifurcated phylogeny with the exception of an unresolved trichotomy within cyprinids (Fig. 2), and most internal branches were supported by 100% PPs. In all 3 data sets, monophyly of the order Cypriniformes, as well as that of the families Catostomidae and Cyprinidae (superfamily Cyprinoidea) were supported by 100% PPs. Monophyly of the superfamily Cobitoidea (Cobitidae + Balitoridae + Catostomidae + Gyrinocheilidae) was reproduced in the Bayesian tree derived from $12_n3_rRT_n$ (RY) coding (Fig. 2), although support for this clade was weak (63%), and the other 2 data sets (123_nRT_n and 12_nRT_n) did not yield a monophyletic Cobitoidea (for details, see below).

Since Fink and Fink (1981, 1996), who corroborated monophyly of the Cypriniformes based on nine morphological synapomorphies, no additional evidence has been presented to support this hypothesis. As far as we know, this study provides the first convincing, independent line of evidence for monophyly of the Cypriniformes based on whole mt genome sequences from all 5 currently recognized families.

It should be noted that cobitine members of the family Cobitidae (*Cobitis*, *Misgurnus*, *Pangio*, *Acanthopsis*) were more closely related to core members of the family Balitoridae, with exclusion of the enigmatic *Vaillantella*, than to the remaining Cobitidae in the subfamily Botinae (*Leptobotia*, *Botia*), rendering the family paraphyletic. The Balitoridae was also paraphyletic because of the unstable positions of *Vaillantella* (indicated by an asterisk in Fig. 2) in the resulting trees. None of the Bayesian trees were compatible with the monophyly of each of the 2 families in any of the pooled 10,800 trees from the 3 data sets, indicating that probabilities of their monophyly were $< 1/10{,}800$ ($p < 0.001$) in the Bayesian context. The 2 loach families (Cobitidae + Balitoridae), however, together formed a monophyletic group with a strong statistical support of 100% PPs in all the 3 data sets. Thus, it appears that future taxonomic revisions will be necessary after additional taxon sampling across these diverse families to accommodate their nonmonophyletic nature discovered from whole mt genome sequences.

For shallower relationships within each of the above 3 major clades (Catostomidae, Cobitidae + Balitoridae, and Cyprinidae), the 3 data sets were largely congruent with the exception of 4 internal branches with relatively low (61% to 93%) and high (100%) PPs (asterisks in Fig. 2). For deeper relationships among the 4 major clades (the above 3 clades + Gyrinocheilidae, represented by a single species *Gyrinocheilus aymonieri*), topologies derived

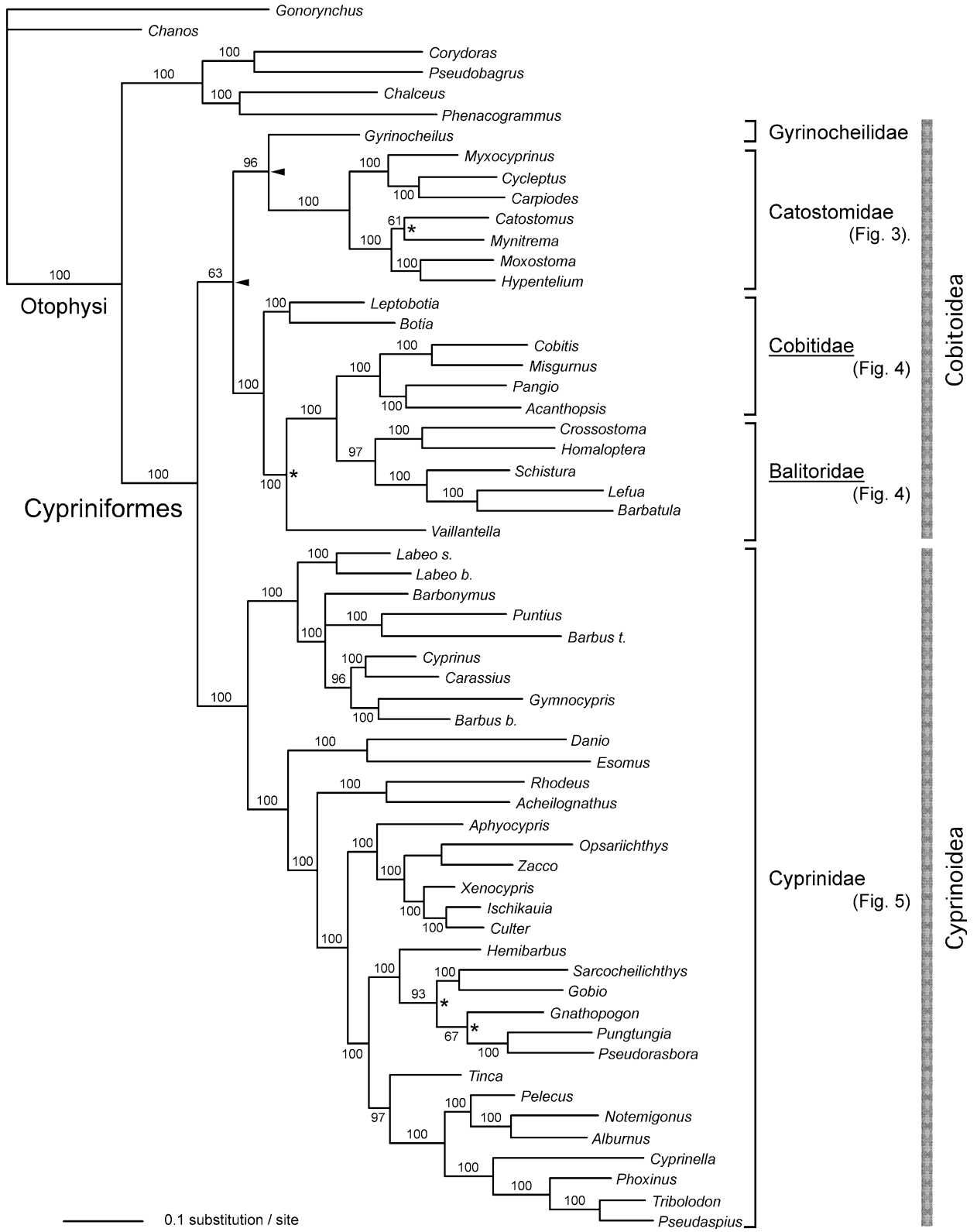

Fig. 2. The 50% majority rule consensus tree of the 10,800 pooled trees from the 4 independent Bayesian analyses using RY ($12_n3_rRT_n$) coding. Numerals beside internal branches indicate Bayesian posterior probabilities (shown as percentages). Topological incongruities among the 3 data sets are denoted by asterisks and arrowheads for shallower and deeper relationships, respectively. Limits of the superfamily Cobitoidea (sensu Sawada 1982) and Cyprinoidea (= Cyprinidae) are indicated. Note that the Cobitidae and Balitoridae families are reciprocally paraphyletic (underlined), although the 2 loach families together form a strongly supported monophyletic group.

Table 4. Statistical comparisons among alternative tree topologies derived in this study of 53 cypriniform taxa used in this study using KH and S-H tests

Tree[a]	−ln L	Diff −ln L	p (KH test)[b]	p (SH test)[b]
123_nRT_n				
1	308422.353	(Best)	–	–
2	308451.718	29.364	0.137	0.155
3	308462.145	39.791	0.090	0.063
$12_n3_rRT_n$				
1	268708.518	55.360	0.131	0.215
2	268653.157	(Best)	–	–
3	268690.740	37.583	0.214	0.352
12_nRT_n				
1	115133.731	36.241	0.068	0.045*
2	115109.133	15.433	0.416	0.388
3	115097.490	(Best)	–	–

[a] Trees nos. 1, 2, and 3 correspond to 50% majority consensus trees derived from 123_nRT_n, $12_n3_rRT_n$, and 12_nRT_n coding, respectively. For differences in relationships among alternative trees, see Table 5.
[b] * $p < 0.05$.

from the 3 data sets greatly differed, with 2 clades indicated by arrowheads (Fig. 2) collapsed in a strict consensus tree. Statistical differences between any pairs of the 3 topologies from the 3 data sets were, however, marginal in both KH ($p = 0.068$ to 0.461) and SH ($p = 0.045$ to 0.388) tests (Table 4), indicating that these alternative topologies could not be rejected statistically .

The relationships among the 4 major clades derived from the 3 data sets differ from either of the 2 previously proposed hypotheses based on morphologies (Wu 1981; Siebert 1987). Neither the hypothesis by Wu (1981), who proposed that the loach family Balitoridae was more closely related to the Cyprinidae than to another loach family Cobitidae, nor a tree compatible with Siebert's (1987) hypothesis, ((((Cobitidae + Balitoridae) Catostomidae) Gyrinocheilidae) Cyprinidae) was found in any of the pooled 10,800 trees from the 3 data sets, indicating that the probability of these hypotheses is $< 1/10{,}800$ ($p < 0.001$) in the Bayesian context.

Source of Incongruence

To explore a source of the incongruities among different codings (123_nRT_n vs. $12_n3_rRT_n$ vs. 12_nRT_n), we examined tree topologies yielded by the Bayesian analyses for each of the 3 data sets and 5 partitions (the third codon positions treated in 2 different manners) with special reference to relationships among the previously mentioned 4 major clades (Catostomidae, Cobitidae + Balitoridae, Cyprinidae, and Gyrinocheilidae). Table 5 lists frequency distributions of the sampled trees compatible with all possible 15 topologies among the 4 major clades and that are incompatible with such criteria (e.g., nonmonophyly for each of the 4 major clades).

It appears that the partitioned Bayesian analyses with the 3 different data sets have yielded different sets of trees (Table 5). Although sampled trees from the 3 data sets greatly overlapped each other, those trees with the highest frequencies for 123_nRT_n, $12_n3_rRT_n$, and 12_nRT_n codings were (((MS)A)L) (M, minnows = Cypriniformes; S, suckers = Catostomidae; A, algae eater = Gyrinocheilidae; and L, loaches = Cobitidae + Balitoridae), (((AS)L)M), and (((AL)M)S), respectively, none of which have common clades shared among the data sets (Table 5).

Interestingly, 2 modes of the topology from analyses of 1_n and 3_r partitions were identical and agreed with that derived from $12_n3_rRT_n$ (RY) coding, whereas a mode of the topology from analyses of 3_n partition agreed with that derived from 123_nRT_n coding (modes are indicated by boldface type in Table 5). It appears that inclusion of the third codon–position transitions (3_n) greatly affected tree topology searches through MCMC samplings for such ambiguous relationships and that the optimum topology has shifted considerably from (((MS)A)L) to (((AS)L)M) by simply reducing them into RY (3_r). Also, it should be noted that the remaining 3 partitions were either noninformative (2_n almost unanimously supported trees incompatible with the possible 15 topologies) or less informative (R_n and T_n were unable to support any specific topologies; Table 5) regarding basal relationships (Table 5), strongly suggesting a lack of historical signals in these partitions despite their higher signal-to-noise ratios (treeness/RCV = 13.647 to 37.759; Table 3) than those of third codon positions (treeness/RCV = 1.806 and 2.120 in 3_n and 3_r, respectively; Table 3).

Considering quickly saturated transitional changes in the third codon positions (Fig. 1), the lowest signal-to-noise ratio in 3_n partition (Table 3), and highly biased sampled trees derived from the analyses of 3_n partition (Table 5), it is likely that noise has been accumulated systematically (not randomly) through transitional multiple hits in the third codon positions, which may eventually lead to erroneous estimations of phylogenies regarding these basal relationships. Alternatively, a lack of historical signals in 2_n, R_n, and T_n partitions for these basal relationships is evident when comparing frequency distributions of the sampled trees (Table 5). We therefore conclude that RY-coding, which takes only transversions into account, effectively removes the likely noises from the third codon positions and avoids a lack of signal by retaining all available positions in the data set. This conclusion is in agreement with that reached by Phillips and Penny (2003) for monotremes, by Delsuc et al. (2003) for hexapods, and by Harrison et al. (2004) for avian mt genomes. In the following

Table 5. Number of trees compatible with 15 possible topologies and incompatible with those topologies among the 4 major clades

Topology	123_nRT_n	$12_n3_rRT_n$	12_nRT_n	1_n	2_n	3_n	3_r	R_n	T_n
(((MS)A)L)	**10382 (96.1)**	18 (0.2)	195 (1.8)	8 (0.4)	0 (0.0)	**1796 (99.7)**	0 (0.0)	37 (2.1)	162 (9.0)
(((AS)L)M)	280 (2.6)	**6425 (59.5)**	745 (6.9)	**840 (46.7)**	0 (0.0)	3 (0.2)	**988 (54.9)**	59 (3.3)	181 (10.1)
(((AS)M)L)	79 (0.7)	3971 (36.8)	842 (7.8)	7 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	41 (2.3)	**389 (21.6)**
(((AL)S)M)	51 (0.5)	339 (3.1)	1411 (13.1)	118 (6.6)	3 (0.2)	0 (0.0)	35 (1.9)	133 (7.4)	69 (3.8)
((MS)(AL))	8 (0.1)	20 (0.2)	1787 (16.5)	252 (14.0)	14 (0.8)	1 (0.1)	0 (0.0)	47 (2.6)	43 (2.4)
((ML)(SA))	0 (0.0)	18 (0.2)	120 (1.1)	28 (1.6)	0 (0.0)	0 (0.0)	0 (0.0)	33 (1.8)	239 (13.3)
(((AL)M)S)	0 (0.0)	9 (0.1)	**5510 (51.0)**	186 (10.3)	10 (0.6)	0 (0.0)	0 (0.0)	246 (13.7)	43 (2.4)
(((ML)A)S)	0 (0.0)	0 (0.0)	181 (1.7)	15 (0.8)	0 (0.0)	0 (0.0)	0 (0.0)	230 (12.8)	128 (7.1)
(((MS)L)A)	0 (0.0)	0 (0.0)	5 (0.0)	56 (3.1)	0 (0.0)	0 (0.0)	0 (0.0)	103 (5.7)	55 (3.1)
(((MA)L)S)	0 (0.0)	0 (0.0)	2 (0.0)	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	**324 (18.0)**	10 (0.6)
(((ML)S)A)	0 (0.0)	0 (0.0)	2 (0.0)	4 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	133 (7.4)	207 (11.5)
(((MA)S)L)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	87 (4.8)	30 (1.7)
((MA)(SL))	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	94 (5.2)	5 (0.3)
(((SL)A)M)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	12 (0.7)	100 (5.6)	14 (0.8)
(((SL)M)A)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	133 (7.4)	16 (0.9)
Others	0 (0.0)	0 (0.0)	0 (0.0)	285 (15.8)	**1773 (98.4)**	0 (0.0)	765 (42.5)	0 (0.0)	209 (11.6)

[a] (M, minnows = Cyprinidae; S, suckers = Catostomidae; A, algae eater, Gyrinocheilidae; and L, loaches = Cobitidae + Balitoridae).
[b] Sampled by Bayesian analyses for the 3 data sets and 5 partitions used in this study.
Boldface denotes maximum number of topologies sampled for each dataset and partition.

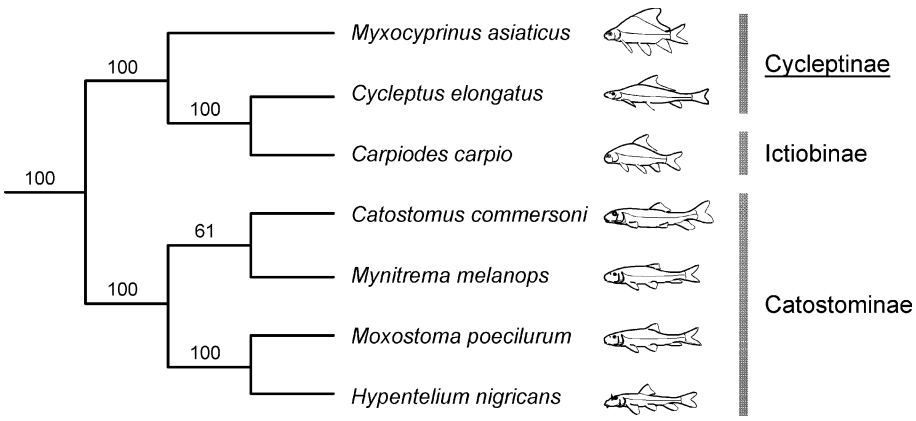

Fig. 3. Catostomid phylogeny derived from $12_n3_rRT_n$ coding. Numbers beside internal branches indicate Bayesian posterior probabilities (shown as percentages). Subfamilial classification of Smith (1992) is indicated, and the resulting nonmonophyletic subfamily is underlined.

presentations, we use the Bayesian tree derived from $12_n3_rRT_n$ (RY) coding (Fig. 2) as the best phylogenetic estimate for the 53 cypriniforms.

Catostomid Phylogeny

Smith (1992) provided the first comprehensive analysis of catostomid relationships based on 64 taxa and 157 morphological, biochemical, and early life history characters and used that hypothesis to classify the catostomid genera into 3 subfamilies (Ictiobinae, Cycleptinae, and Catostominae). Subsequently, Harris and Mayden (2001) analyzed catostomid relationships based on partial sequences of the mitochondrial 12S and 16S rRNA genes from 16 catostomid species. Their trees reproduced monophyly of the family Catostomidae and each of the 2 subfamilies (Ictiobinae and Catostominae), although the remaining subfamily (Cycleptinae) was paraphyletic or polyphyletic. Our mt genome tree (Fig. 3) supported the nonmonophyly of the Cycleptinae, as one of the cycleptines (*Cycleptus elongatus*) was more closely related to an ictiobine (*Carpiodes carpio*) than to another cycleptine (*Myxocyprinus asiaticus*), the only Asian member of the family Catostomidae. There was no tree compatible with monophyly of the Cycleptinae in any of the pooled 10,800 trees from the 3 data sets, indicating that probability of this hypothesis is $< 1/10{,}800$ ($p < 0.001$) in the Bayesian context. Of course, more extensive taxon sampling is necessary before re-evaluation of this putatively paraphyletic subfamily.

Cobitid + Balitorid Phylogeny

Within the monophyletic group comprising the 2 paraphyletic loach families (Cobitidae + Balitoridae), phylogenetic positions of monotypic *Vaillantella* was the most problematic (Fig. 4). This is consistent with the previous views from 2 cladistic studies of Sawada (1982) and Siebert (1987) based on

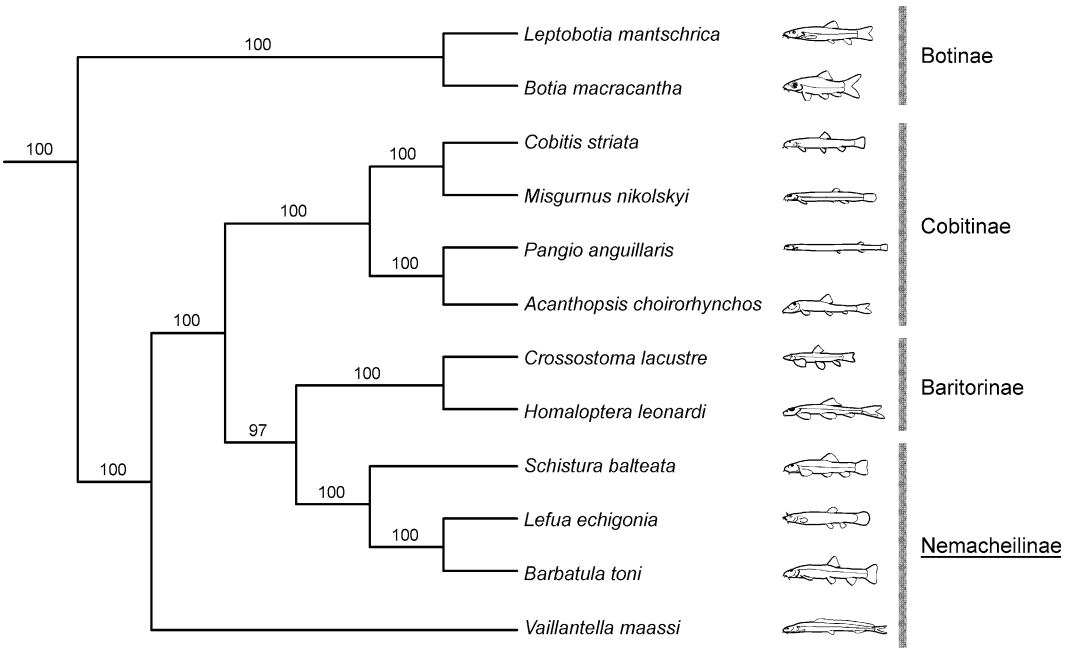

Fig. 4. Cobitid + balitorid phylogenies derived from $12_n3_rRT_n$ coding. Numbers beside internal branches indicate Bayesian posterior probabilities (shown as percentages). Subfamilial classification of Sawada (1982) is indicated, and the resulting nonmonophyletic subfamily is underlined.

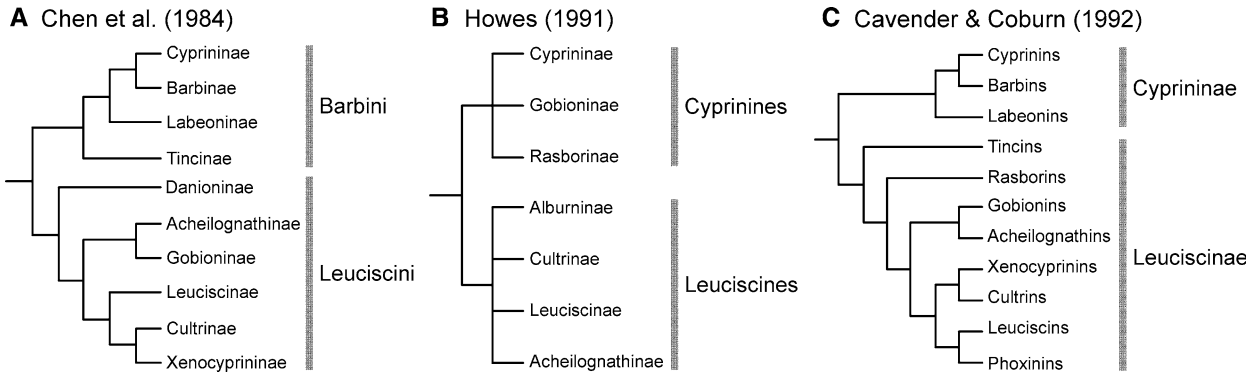

Fig. 5. Three alternative morphology-based hypotheses of relationships among cyprinid subgroups (subfamilies or informal taxa). Note that **A** Chen et al. (1984) and **B** Howes (1991) assigned subfamilies to terminal taxa, whereas **C** Cavender and Coburn (1992) assigned subfamilies to the 2 basal lineages.

morphological data, in which *Vaillantella* was considered to be a member of the 2 different families. With the exclusion of this enigmatic species, each of the currently recognized 4 subfamilies are monophyletic in the present tree (Fig. 4), although their overall relationships differed from the previous morphology-based hypotheses (Sawada 1982; Siebert 1987) in that the Botinae was the basal-most clade sister to a clade comprising other 3 subfamilies, rendering the Cobitidae paraphyletic. No trees were compatible with the monophyly of the Cobitidae in any of the pooled 10,800 trees from the 3 data sets, indicating that probability of this hypothesis is $< 1/10{,}800$ ($p < 0.001$) in the Bayesian context. Because there is another problematic genus *Ellopostoma*, once placed in the family Balitoridae by Siebert (1987) and subsequently moved to the family Cobitidae by Roberts (1989), it is premature to revise the classification of this monophyletic group (Cobitidae + Balitoridae) before more extensive taxon sampling has been conducted in a future study.

Cyprinid Phylogeny

Previous morphological studies have failed to reach a consensus regarding the numbers, constituent members, and monophyly of the cyprinid subfamilies or subgroups (Fig. 5). Based on analysis of 25 morphological characters, Chen et al. (1984) divided cyprinids into 2 series (Barbini and Leuciscini) and recognized 4 and 6 subfamilies within each series, respectively (Fig. 5A). Recognition of these lineages,

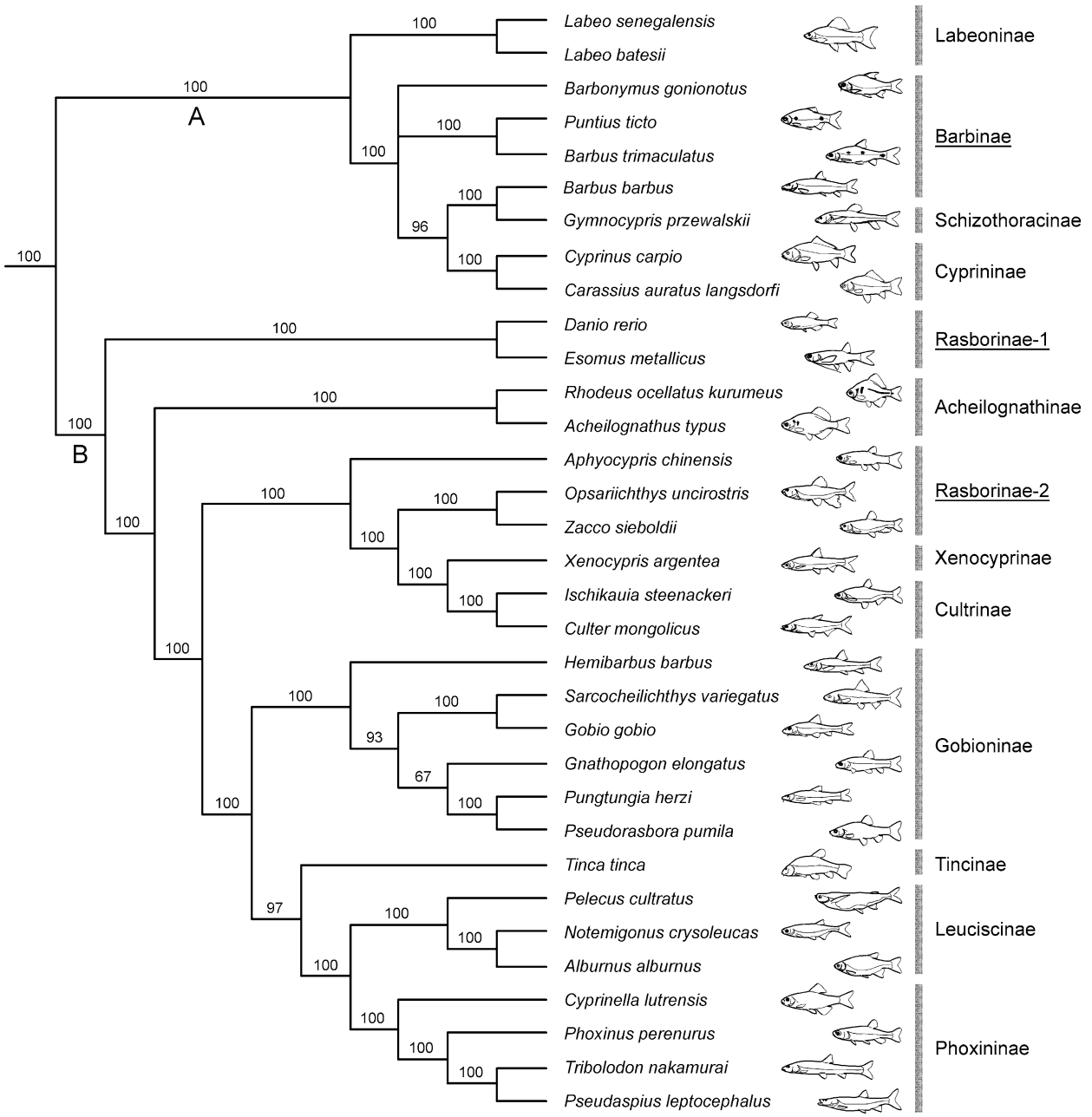

Fig. 6. Cyprinid phylogenies derived from $12_n3_rRT_n$ coding. Numbers beside internal branches indicate Bayesian posterior probabilities (shown as percentages). Subfamilial classification employed in this study (see Table 1) is indicated, and the resulting nonmonophyletic subfamilies are underlined and numbered. Clades A and B correspond to the subfamilies Cyprininae and Leuciscinae in Cavender and Coburn (1992).

however, was based on an autapomorphy and no synapomorphy is given for each sister group (Howes 1991). Subsequently, Howes (1991) proposed a diphyletic arrangement into 2 lineages for the cyprinids (informal groups named cyprinines and leuciscines) based on presence or absence of the anterior maxillary barbel with an accompanying maxillary foramen (Fig. 5B). He recognized 7 subfamilies (plus 5 lineages within the Cyprininae) and provided a synopsis of those subfamilies he considered valid with their diagnostic characters. Although Cavender and Coburn's (1992) hypothesis was similar to that of Chen et al. (1984), it was the first cyprinid phylogeny based on a rigorous cladistic analysis of morphological variation in the order (Fig. 5C). On the basis of their analysis, Cavender and Coburn (1992) assigned 2 subfamilies (Cyprininae and Leuciscinae) to 2 major lineages within the cyprinids and 3 and 8 informal taxa to each subfamily (Fig. 5C), respectively. None of the previous morphology-based hypotheses are in full agreement with the present molecular phylogeny (Fig. 6), although the basal-

most split into 2 major lineages (clades A and B in Fig. 6) is consistent with that of Cavender and Coburn (1992) and supports the recognition of 2 subfamilies, Cyprininae and Leuciscinae.

Numerous molecular studies have been conducted for resolving intrarelationships of the Cyprinidae and addressing validity of its subfamilies. Taxon sampling in most of these studies, however, was restricted either to a certain geographic region, such as Europe (Briolay et al. 1998; Gilles et al. 1998, 2001; Zardoya and Doadrio 1998, 1999a,b), the Middle East (Durand et al. 2002), or East Asia (Liu and Chen 2003; He et al. 2004b), or to specific taxa in a specific geographic region, such as the Acheilognathinae in East Asia (Okazaki et al. 2001), or Labeoninae (Li et al. 2005), Schizothoracinae (He et al. 2004a), and Xenocyprininae (Xiao et al. 2001) in China, or Phoxininae in North America (Simons et al. 2003). Because of the limited geographic representation of taxon sampling in these studies, there has been no large-scale molecular phylogenetic evaluation of the family Cyprinidae until recently.

Exception to these limited analyses is a study of Cunha et al. (2002), who compiled the published 73 cyt *b* gene sequences from various studies and added 16 newly determined sequences to construct a large data matrix including 33 European, 38 Asian, and 18 North American taxa (total = 89 spp.). Because of the short sequence lengths (1140 bp) and probably because of analytic methods employed (neighbor-joining analysis based on Jukes-and-Cantor distance and unweighted maximum-parsimony analysis), phylogenies from this study were not resolved well enough to draw explicit conclusions regarding the basal relationships within the family, although they identified 5 major lineages within the Cyprinidae.

The first lineage Cunha et al. (2002) recognized was a clade comprising European leuciscines (including *Notemigonus*) + North American phoxines, which appears to correspond to that of Leuciscinae + Phoxininae in the present tree (Fig. 6). Some members of the second lineage (European gobionines + *Pseudorasbora*) and third lineage (primarily Asian groups, cultrines + acheilognathines + other gobionines + xenocyprinines) in Cunha et al. (2002) were found, in the present phylogeny (Fig. 6), in reciprocally paraphyletic clades consisting the Gobioninae + Cultrinae + Xenocyprininae + Rasborinae + Acheilognathinae, although congruence between those clades from the 2 studies was ambiguous because of different taxon sampling. None of the species from the fourth lineage (*Abbotina* + *Sinocyclocheilus* + *Acrossocheilus*) was sampled in the present study and the fifth lineage (cyprinines + barbines + labeonines) in Cunha et al. (2002) was in full agreement with clade A (Fig. 6).

Given the limited taxon sampling from the cyprinids (33 species from >2500 currently recognized species; Table 1), the present study represents the first example to clearly identify the 2 major lineages within cyprinids based on the molecular data (clades A and B; Fig. 6), which are congruent with the 2 subfamilies (Cyprininae and Leuciscinae) recognized by Cavender and Coburn (1992) based on morphologies. In addition, the present study clearly resolves the basal relationships within each major lineage. In clade A, labeonines were placed at the basal-most position, sister to a clade comprising a paraphyletic barbinines, a schizothoracine, and cyprinines (Fig. 6). Intrarelationships within the latter clade, however, were ambiguous because of an unresolved trichotomy among some lineages. In clade B, 2 South Asian rasborines (*Danio* and *Esomus*) were placed at the basal-most position and sister to other members of the clade. Nested sister-group relationships among clades were sequentially represented by the following 6 major clades: (1) acheilognathines, (2) a clade consisting of East Asian rasborines, a xenocyprinine, and cultrines, (3) gobionines, (4) tincine, (5) leuciscines, and (6) phoxinines (Fig. 6). With our current lack of African and/or south Asian asborines, cultrines, and leuciscines, as well as some Asian endemic cyprinids (e.g., *Psilorhynchus*), we are unable at this time to evaluate the validity of these clades with confidence.

Concluding Remarks

Resolution of higher-level relationships of cypriniforms is intrinsically difficult because there has been neither a good reference taxonomy nor a comprehensive phylogenetic study that encompasses the whole spectrum of cypriniform diversity. Indeed, there has been no consensus about familial and subfamilial classifications of the cypriniforms. We had to construct a mosaic classification of the cypriniforms from different investigators (Table 1) to express their tremendous species diversity and to justify our taxon sampling strategy. Given these difficulties, however, the present study represents an excellent initiative that provides a testable framework of phylogenetic history that is critical to galvanize future studies. Continuing efforts can focus on more extensive taxon sampling that will lead to a more correct estimation of the evolutionary history and biogeography of the tremendous Cypriniformes diversity and more robust evolutionary inferences from this extraordinarily diverse and biotically complex clade of fishes that is extremely important in comparative biological studies.

Acknowledgments. This study has been conducted as a part of the Cypriniformes Tree of Life (CToL) Project (http://www.cypriniformes.org). For donation of the study materials, we sincerely thank T. Asahida, H. L. Bart Jr., J. Bohlen, T. Eros, Y. Fujimoto, H. Imai, Y. P. Kartavtsev, K. Katsura, S. Lavoué, N. E. Mandrak, M. Matsuda, T. Mukai, J. Nakajima, H. Persat, P. Rab, V. V. Sviridov, K. Takahashi, T. Ueda, T. Urano, A. A. Varaksin, E. O. Wiley, and T. Yokoyama. We also thank Y. Kumazawa for advice on data analysis. Critical comments from 2 anonymous reviewers were helpful to improve the manuscript. This study was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and Japan Science for the Promotion of Science (Grants No. 12NP0201, 13556028, 15570090, 15380131, and 17207007).

References

Adachi J, Hasegawa M (1996) Model of amino acid substitution in proteins encoded by mitochondrial DNA. J Mol Evol 42:459–468

Alves MJ, Collares-Pereira MJ, Dowling TE, Coelho MM (2002) The genetics of maintenance of an all-male lineage in the *Squalius alburnoides* complex. J Fish Biol 60:649–662

Arai R (1982) A chromosome study on two cyprinid fishes *Acrossocheilus labiatus* and *Psuedorasbora pumila pumila*, with notes on Eurasian cyprinids and their karyotypes. Bull Nat Sci Mus Tokyo Ser A (Zool) 8:131–152

Asahida T, Kobayashi T, Saitoh K, Nakayama I (1996) Tissue preservation and total DNA extraction from fish stored at ambient temperature using buffers containing high concentration of urea. Fish Sci 62:727–730

Baird IG, Inthaphaisy V, Kisouvannalath P, Phylavanh B, Mounsouphom B (1999) The fishes of southern Lao. Lao Community Fisheries and Dolphin Protection Project. Ministry of Agriculture and Forestry, Lao PDR, 161 pp

Berra T (2001) Freshwater fish distribution. Academic, London, UK

Briolay J, Galtier N, Brito RM, Bouvet Y (1998) Molecular phylogeny of Cyprinidae inferred from cytochrome *b* DNA sequences. Mol Phylogenet Evol 9:100–108

Broughton RE, Milam JE, Roe BA (2001) The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Res 11:1958–1967

Cavender TM, Coburn MM (1992) Phylogenetic relationships of North American Cyprinidae. In: Mayden RL (ed) Systematics historical ecology and North American freshwater fishes. Stanford University Press, Palo Alto, CA, pp 293–327

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

Chen X-L, Yue P-Q, Lin R-D (1984) Major groups within the family Cyprinidae and their phylogenetic relationships. Acta Zootaxon Sinica 9:424–440

Chen Y-Y et al. (1997) Fauna Sinica. Osteichthyes: Cypriniformes II. Science Press, Beijing, China

Cheng S, Chang S-Y, Gravitt P, Respess R (1994) Long PCR. Nature 369:684–685

Clements MD, Bart HL Jr, Herley DL (2004) Isolation and characterization of two distinct growth hormone cDNAs from the tetraploid smallmouth buffalofish (*Ictiobus bubalus*). Gen Comp Endocrinol 136:411–418

Coburn MM, Cavender TM (1992) Interrelationships of North American cyprinid fishes. In: Mayden RL (ed) Systematics historical ecology and North American freshwater fishes. University Press, Palo Alto, CA, pp 328–373

Cunha C, Mesquita N, Dowling TE, Gilles A, Coelho MM (2002) Phylogenetic relationships of Eurasian and American cyprinids using cytochrome *b* sequences. J Fish Biol 61:929–944

Delsuc F, Phillips MJ, Penny D (2003) Technical comment on "Hexapod origins: Monophyletic or paraphyletic?" Science 301:1482

Dettai A, Lecointre G (2005) Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. CR Biol 328:674–689

De Rijk P, Caers A, Van de Peer Y, De Wachter R (1998) Database of the structure of large ribosomal subunit RNA. Nucleic Acids Res 26:183–186

Durand J-D, Tsigenopoulos CS, Ünlü E, Berrebi P (2002) Phylogeny and biogeography of the family Cyprinidae in the Middle East inferred from cytochrome *b* DNA: Evolutionary significance of this region. Mol Phylogenet Evol 22:91–100

Duyk G, Schmitt K (2001) Fish x 3. Nat Genet 27:8–9

Fink SV, Fink WL (1981) Interrelationships of the ostariophysan fishes. Zool J Linn Soc 72:297–353

Fink SV, Fink WL (1996) Interrelationships of ostariophysan fishes. In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fishes. Academic, San Diego, CA, pp 209–249

Gilles A, Lecointre G, Faure E, Chappaz R, Brun G (1998) Mitochondrial phylogeny of the European cyprinids: implications for their systematics, reticulate evolution, and colonization time. Mol Phylogenet Evol 19:22–33

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

Hänfling B, Brandl R (2000) Phylogenetics of European cyprinids: Insights from allozymes. J Fish Biol 57:265–276

Harris PM, Mayden RL (2001) Phylogenetic relationships of major clades of Catostomidae (Teleostei: Cypriniformes) as inferred from mitochondrial SSU and LSU rDNA sequences. Mol Phylogenet Evol 20:225–237

Harrison GL, McLenachan PA, Phillips MJ, Slack KE, Cooper A, Penny D (2004) Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. Mol Biol Evol 21:974–983

He D, Chen Y, Chen Y, Chen Z (2004a) Molecular phylogeny of the specialized schizothoracine fishes (Teleostei: Cyprinidae), with their implications for the uplift of the Qinghai-Tibetan Plateau. Chinese Sci Bull 49:39–48

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

Hillis DM (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst Biol 47:3–8

Howes GJ (1991) Systematics and biogeography: An overview. In: Winfield IJ, Nelson JS (eds) Cyprinid fishes: Systematics biology and exploitation. Chapman & Hall, London, UK, pp 1–33

Inoue JG, Miya M, Tsukamoto K, Nishida M (2000) Complete mitochondrial DNA sequence of the Japanese sardine, *Sardinops melanostictus*. Fish Sci 66:924–932

Inoue JG, Miya M, Aoyama J, Ishikawa S, Tsukamoto K, Nishida M (2001a) Complete mitochondrial DNA sequence of the Japanese eel, *Anguilla japonica*. Fish Sci 67:118–125

Inoue JG, Miya M, Tsukamoto K, Nishida M (2001b) Complete mitochondrial DNA sequence of *Conger myriaster* (Teleostei: Anguilliformes): Novel gene order for vertebrate mitochondrial genomes and the phylogenetic implications for anguilliform families. J Mol Evol 52:311–320

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

Inoue JG, Miya M, Tsukamoto K, Nishida M (2001d) Complete mitochondrial DNA sequence of the Japanese anchovy, *Engraulis japonicus*. Fish Sci 67:828–835

Inoue JG, Miya M, Tsukamoto K, Nishida M (2003) Evolution of the deep-sea gulper eel mitochondrial genomes: Large-scale gene rearrangements originated within the eels. Mol Biol Evol 20:1911–1917

Inoue JG, Miya M, Tsukamoto K, Nishida M (2004) Mitogenomic evidence for the monophyly of elopomorph fishes (Teleostei) and the evolutionary origin of the leptocephalus larva. Mol Phylogenet Evol 32:274–286

Ishiguro NB, Miya M, Nishida M (2001) Complete mitochondrial DNA sequence of ayu, *Plecoglossus altivelis*. Fish Sci 67:474–481

Ishiguro NB, Miya M, Nishida M (2003) Basal euteleostean relationships: A mitogenomic perspective on the phylogenetic reality of the ''Protacanthopterygii.'' Mol Phylogenet Evol 27:476–488

Ishiguro NB, Miya M, Inoue JG, Nishida M (2005) *Sundasalanx* (Sundasalangidae) is a progenetic clupeiform, not a closely-related group of salangids (Osmeriformes): Mitogenomic evidence. J Fish Biol 67:561–569

Johnson GD, Patterson C (1993) Percomorph phylogeny: A survey of acanthomorphs and a new proposal. Bull Mar Sci 52:554–626

Kawaguchi A, Miya M, Nishida M (2001) Complete mitochondrial DNA sequence of *Aulopus japonicus* (Teleostei: Aulopiformes), a basal Eurypterygii: Longer DNA sequences and higher-level relationships. Ichthyol Res 48:213–223

Kimmel CB (1989) Genetics and early development of zebrafish. Trends Genet 5:283–288

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol 29:170–179

Kottelat M, Britz R, Hui TH, Witte KE (2005) *Paedocypris*, a new genus of Southeast Asian cyprinid fish with a remarkable sexual dimorphism, comprises the world's smallest vertebrate. Proc R Soc Lond B Biol Sci 273:895–899

Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J Mol Evol 37:380–398

Kumazawa Y, Azuma Y, Nishida M (2004) Tempo of mitochondrial gene evolution: Can mitochondrial DNA be used to date old divergence? Endocytobiosis Cell Res 15:136–142

Lanyon S (1988) The stochastic mode of molecular evolution: What consequences for systematic investigators? Auk 105:565–573

Larget B, Simon DL (1999) Markov Chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol Biol Evol 16:750–759

Lavoué S, Miya M, Inoue JG, Saitoh K, Ishiguro NB, Nishida M (2005) Molecular systematics of the gonorynchiform fishes (Teleostei) based on whole mitochondrial genome sequences: Implications for higher-level relationships within the Otocephala. Mol Phylogenet Evol 37:165–177

Lê HLV, Lecointre G, Perasso R (1993) A 28S rRNA based phylogeny of the gnathostomes: First steps in the analysis of conflict and congruence with morphologically based cladograms. Mol Phylogenet Evol 2:31–51

Leggatt RA, Iwama GK (2003) Occurrence of polyploidy in the fishes. Rev Fish Biol Fish 13:237–246

Li J, Wan X, He S, Chen Y (2005) Phylogenetic studies of Chinese labeonine fishes (Teleostei: Cyprinidae) based on the mitochondrial 16S rRNA gene. Prog Nat Sci 15:213–219

Liu H, Chen Y (2003) Phylogeny of the East Asian cyprinids inferred from sequences of the mitochondrial DNA control region. Can J Zool 81:1938–1946

Lôpez JA, Chen W-J, Ortí G (2004) Esociform phylogeny. Copeia 2004:449–464

Mabuchi K, Miya M, Senou H, Suzuki T, Nishida M (2006) Complete mitochondrial DNA sequence of the Lake Biwa wild strain of common carp (*Cyprinus carpio* L.): Further evidence for an ancient origin. Aquaculture 257:68–77

Martin AP, Burg TM (2002) Perils of paralogy: Using HSP70 genes for inferring organismal phylogenies. Syst Biol 51:570–587

Mau B, Newton MA, Larget B (1999) Bayesian phylogenetic inference via Markov Chain Monte Carlo methods. Biometrics 55:1–12

Miller RE, Buckley TR, Manos PS (2002) An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. Syst Biol 51:740–753

Minegishi Y, Aoyama J, Inoue JG, Miya M, Nishida M, Tsukamoto K (2005) Molecular phylogeny and evolution of the freshwater eels genus *Anguilla* based on the whole mitochondrial genome sequences. Mol Phylogenet Evol 34:134–146

Mirza RS, Chivers DP (2003) Fathead minnows learn to recognize heterospecific alarm cues they detect in the diet of a known predator. Behaviour 140:1359–1369

Miya M, Nishida M (1999) Organization of the mitochondrial genome of a deep-sea fish *Gonostoma gracile* (Teleostei: Stomiiformes): First example of transfer RNA gene rearrangements in bony fishes. Mar Biotechnol 1:416–426

Miya M, Nishida M (2000) Use of mitogenomic information in teleostean molecular phylogenetics: A tree-based exploration under the maximum-parsimony optimality criterion. Mol Phylogenet Evol 17:437–455

Miya M, Kawaguchi A, Nishida M (2001) Mitogenomic exploration of higher teleostean phylogenies: A case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol Biol Evol 18:1993–2009

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. Mol Phylogenet Evol 26:121–138

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

Miya M, Satoh TP, Nishida M (2005) The phylogenetic position of toadfishes (order Batrachoidiformes) in the higher ray-finned fish as inferred from partitioned Bayesian analysis of 102 whole mitochondrial genome sequences. Biol J Linn Soc 85:289–306

Murakami J, Yamashita Y, Fujitani H (1998) The complete sequence of mitochondrial genome from a gynogenetic triploid ''ginbuna'' (*Carassius auratus longsdorfi*). Zool Sci 15:335–337

Nelson JS (1994) Fishes of the world, third ed. Wiley, NY,

Okazaki M, Naruse K, Shima A, Arai R (2001) Phylogenetic relationships of bitterings based on mitochondrial 12S ribosomal DNA sequences. J Fish Biol 58:89–106

Phillips MJ, Penny D (2003) The root of the mammalian tree inferred from whole mitochondrial genomes. Mol Phylogenet Evol 28:171–185

Rannala B, Yang Z (1996) Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. J Mol Evol 43:304–311

Roberts TR (1989) The freshwater fishes of western Borneo (Kalimantan Barat Indonesia). Mem Calif Acad Sci 14:i–xii, 1–210

Roberts TR, Warren TJ (1994) Observations of fishes and fisheries in southern Laos and northeastern Cambodia October 1993–February 1994. Nat Hist Bull Siam Soc 42:87–115

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Saitoh K, Kim I-S, Lee E-H (2004) Mitochondrial gene introgression between spined loaches via hybridogenesis. Zool Sci 21:795–798

Saitoh K, Miya M, Inoue JG, Ishiguro NB, Nishida M (2003) Mitochondrial genomics of ostariophysan fishes: Perspectives on phylogeny and biogeography. J Mol Evol 56:464–472

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

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116

Siebert DJ (1987) Interrelationships among families of the order Cypriniformes (Teleostei). Doctoral dissertation, The City University of New York, New York, NY,

Simmons MP, Miya M (2004) Efficiently resolving the basal clades of a phylogenetic tree using Bayesian and parsimony approach: A case study using mitogenomic data from 100 higher teleost fishes. Mol Phylogenet Evol 31:351–362

Simons AM, Berentzen P, Mayden RM (2003) Molecular systematics of North American phoxinin genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Zool J Linn Soc 139:63–80

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. Stanford University Press, Palo Alto, CA, pp 778–826

Smith WL, Wheeler WC (2004) Polyphyly of the mail-cheeked fishes (Teleostei: Scorpaeniformes): evidence from mitochondrial and nuclear sequence data. Mol Phylogenet Evol 32:627–646

Swofford DL (2001) PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates, MA

Talwar PK, Jhingran AG (1991) Inland fishes of India and adjacent countries. Volume 1. Balkema, Rotterdam, The Netherlands

Tang Q, Xiong B, Yang X, Liu H (2005) Phylogeny of the East Asian botiine loaches (Cypriniformes Botiidae) inferred from mitochondrial cytochrome *b* gene sequences. Hydrobiologia 544:249–258

Tzeng CS, Hui CF, Shen SC, Huang PC (1992) The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. Nucleic Acids Res 20:4853–4858

Van de Peer Y, Caers A, De Rijk P, De Wachter R (1998) Database on the structure of small ribosomal subunit RNA. Nucleic Acids Res 26:179–182

Vrijenhoek RC, Dawley RM, Cole CJ, Bogart JP (1989) A list of the known unisexual vertebrates. In: Dawley RM, Bogart JP (eds) Evolution and ecology of unisexual vertebrates. Bull 466. New York State Museum, New York, NY, pp 19–23

Wiley EO, Johnson GD, Dimmick WW (2000) The interrelationships of acanthomorph fishes: A total evidence approach using molecular and morphological data. Biochem Syst Ecol 28:319–350

Wu X, Chen Y, Chen X-L, Chen J-X (1981) Classification and phylogenetic interrelationships among families of suborder Cyprinoidei. Sci China 3:369–376

Yamanoue Y, Miya M, Matsuura K, Katoh M, Sakai H, Nishida M (2004) Mitochondrial genomes and phylogeny of the ocean sunfishes (Tetraodontiformes: Molidae). Ichthyol Res 51:269–273

Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J Mol Evol 39:306–314

Yang Z, Rannala B (1997) Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo method. Mol Biol Evol 14:717–724

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

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

Zardoya R, Doadrio I (1998) Phylogenetic relationships of Iberian Cyprinidae: Systematic and biogeographical implications. Proc Roy Soc Lond Ser B 265:1365–1372

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

Zardoya R, Doadrio I (1999b) Phylogenetic relationships of Greek Cyprinidae: Molecular evidence for at least two independent origins of the Greek cyprinid fauna. Mol Phylogenet Evol 13:122–131