Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 39 (2006) 347–357
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

# Comparison of evolutionary rates in the mitochondrial DNA cytochrome *b* gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes)

Qiongying Tang [a,b], Huanzhang Liu [a,*], Richard Mayden [c], Bangxi Xiong [b]

[a] *Institute of Hydrobiology, Chinese Academy of Sciences, Hubei, Wuhan 430072, PR China*
[b] *College of Fishery, Huazhong Agricultural University, Hubei, Wuhan 430070, PR China*
[c] *Department of Biology, Saint Louis University, 3507 Laclede Ave., St. Louis, MO 63103-2010, USA*

Received 6 July 2005; revised 15 August 2005; accepted 18 August 2005
Available online 4 October 2005

**Abstract**

It is widely accepted that mitochondrial DNA (mtDNA) control region evolves faster than protein encoding genes with few exceptions. In the present study, we sequenced the mitochondrial cytochrome *b* gene (cyt *b*) and control region (CR) and compared their rates in 93 specimens representing 67 species of loaches and some related taxa in the Cobitoidea (Order Cypriniformes). The results showed that sequence divergences of the CR were broadly higher than those of the cyt *b* (about 1.83 times). However, in considering only closely related species, CR sequence evolution was slower than that of cyt *b* gene (ratio of CR/cyt *b* is 0.78), a pattern that is found to be very common in Cypriniformes. Combined data of the cyt *b* and CR were used to estimate the phylogenetic relationship of the Cobitoidea by maximum parsimony, neighbor-joining, and Bayesian methods. With *Cyprinus carpio* and *Danio rerio* as outgroups, three analyses identified the same four lineages representing four subfamilies of loaches, with Botiinae on the basal-most clade. The phylogenetic relationship of the Cobitoidea was ((Catostomidae + Gyrinocheilidae) + (Botiinae + (Balitorinae + (Cobitinae + Nemacheilinae)))), which indicated that Sawada's Cobitidae (including Cobitinae and Botiinae) was not monophyletic. Our molecular phylogenetic analyses are in very close agreement with the phylogenetic results based on the morphological data proposed by Nalbant and Bianco, wherein these four subfamilies were elevated to the family level as Botiidae, Balitoridae, Cobitidae, and Nemacheilidae.


*Keywords:* Cytochrome *b*; Control region; Sequence divergence; Phylogenetic analysis; Cobitoidea

## 1. Introduction

Mitochondrial DNA (mtDNA) sequences, especially the cytochrome *b* (cyt *b*) gene and the control region (CR) are frequently utilized for population genetic and phylogenetic studies of fishes (Liu and Chen, 2003; Moum and Árnason, 2001; Peng et al., 2004; Perdices et al., 2004). The cytochrome *b* gene encodes a protein and evolves relatively slowly, whereas the non-coding CR in vertebrates, presumably because of the lack of coding constraints, evolves rapidly. Sequence variation in the CR consists not only of substitutions but also of indels of various lengths and of variation in number of copies of tandem repeats (Sbisà et al., 1997). Control region, especially the $tRNA^{pro}$ end, has been suggested to have one of the highest substitution rates of all the mitochondrial genes (Brown, 1985; Meyer, 1993). Mutation rate of the CR can be two to five times higher than that of mitochondrial protein-coding genes (Meyer, 1993). However, several reports have challenged the generality of this observation, especially in different fish groups. In rat and mouse, Brown et al. (1986) found a slower rate of substitution in CR than that of protein-coding genes. A slower rate of substitution in CR was also found in salmonid fishes (Bernatchez and Danzmann, 1993; Shedlock et al., 1992), and butterflies of the

* Corresponding author. Fax: +86 27 68780123.
*E-mail address:* hzliu@ihb.ac.cn (H. Liu).

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.08.007

genus *Jalmenus* (Taylor et al., 1993). Zhu et al. (1994) compared relative rates and patterns of sequence evolution in CR and cyt *b* sequences from different populations and species of freshwater rainbow fishes of the genus *Melanotaenia*, and discovered that the overall levels of divergence were similar for these two gene segments but patterns of sequence evolution varied. Crochet and Desmarais (2000) provided evidence for a lower-than-expected interspecific divergence among CRs of gulls and proposed that the slow rate of evolution of CR part III of the gulls could be partly explained by the existence of secondary structures. All these and other studies have been confined to species or genera. Comparison at different levels including species, genera, families, and for genealogical patterns of molecular evolution of these important genera is needed.

Fishes of the family Cobitidae are part of a major lineage of the order Cypriniformes, which is the largest group of freshwater fishes in the world. Presently, five families (Gyrinocheilidae, Catostomidae, Cobitidae, Balitoridae, and Cyprinidae) are recognized as valid in Cypriniformes (Nelson, 1994). However, their phylogenetic relationships remain controversial. Two main hypotheses had been proposed by Wu et al. (1981) and Siebert (1987) (Fig. 1). Wu et al. (1981) suggested that the Balitoridae (=Homalopteridae) was closest to the Cyprinidae and the other families form another monophyletic group. Siebert (1987) proposed that the Cyprinidae forms a single monophyletic group and the non-cyprinid cypriniforms form another monophyletic group, a conclusion supported by some recent investigations (He et al., 1997; Liu et al., 2002). The fact that Gyrinocheilidae and Catostomidae form their own monophyletic group has also been accepted widely. Therefore, the relationship between Cobitidae and Balitoridae and their relationship to other families is the key to resolve the phylogenetic relationship of the whole Cypriniformes.

Regan (1911) first defined the group Cobitidae and divided the family into the subfamilies Cobitinae and Nemacheilinae. Hora (1932) classified the family Homalopteridae, an apparent clade that has been replaced by name as Balitoridae (Kottelat, 1988), into two subfamilies Gastromyzoninae and Homalopterinae (=Balitorinae), and considered the former as a derivative of the Cobitidae with the latter as a descendant of the family Cyprinidae. Berg (1940) divided the Cobitidae into three subfamilies, Botiinae, Cobitinae, and Nemacheilinae, a change that was accepted by many authors at that time (Chen and Zhu, 1984; Nalbant, 1963; Ramaswami, 1953; Wu et al., 1981). After examining 52 characters of 48 species or subspecies, Sawada (1982) transferred the subfamily Nemacheilinae from the family Cobitidae to the family Balitoridae, these two clades form a monophyletic group, the superfamily Cobitoidea. The former group is differentiated into two monophyletic groups Botiinae and Cobitinae which are considered sister groups, and the latter consists of Nemacheilinae and Balitorinae. This classification has been widely accepted (Kottelat, 2001; Nelson, 1994; Siebert, 1987). However, based on molecular phylogenetic analysis of the Cypriniformes, Liu et al. (2002) proposed that the relationships within the Cobitoidei are: Catostomidae + (Gyrinocheilidae + (Botiinae + (Balitoridae + (Cobitinae + Nemacheilinae)))). Thus, the Botiinae forms the basal group to other loaches, a conclusion in general agreement with Nalbant (1963). Furthermore, Nalbant (2002) treated the Botiinae, Cobitinae, and Nemacheiliane as three valid families Botiidae, Cobitidae, and Nemacheilidae. The analyses by Liu et al. (2002) included only a few loach species, precluding an adequate test of the phylogenetic relationship of loaches.

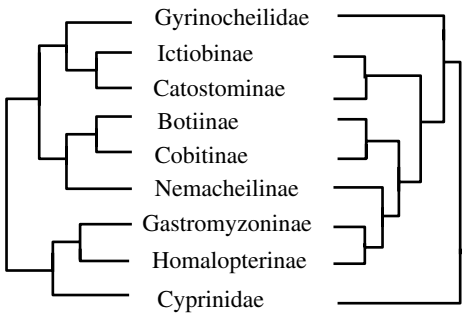


Fig. 1. Two hypotheses concerning the phylogeny of the Cypriniformes. The one on the right was proposed by Siebert (1987), but has been modified slightly. The one on the left is mainly from Wu et al. (1981), with the relationships of the Cobitidae from Chen and Zhu (1984).

In the present study, we sequenced mitochondrial cytochrome *b* gene and CR of the so-called loaches (including the families Cobitidae and Balitoridae) to compare the evolutionary rate of these two segments at different classification levels that has been examined previously, and study the phylogenetic relationship of the Cobitoidea.

## 2. Materials and methods

### 2.1. Samples and DNA extraction

In present study, 93 specimens representing 67 species of loaches and some related taxa in the Cobitoidea were selected for analysis. Two sequences of *Myxocyprinus asiaticus* were obtained from GenBank [AF036176 (cyt *b*), AY017140 (CR)]. Detailed information of specimens is listed in Table 1. The cyt *b* and CR sequences of *Cyprinus carpio* and *Danio rerio* were used as outgroups (NC001606 and NC002333). Muscles from alcohol fixed museum specimens were used for DNA extraction. All specimens belong to the Institute of Hydrobiology, Chinese Academy of Sciences. Total DNA was extracted using standard proteinase K digestion followed by phenol/chloroform extraction (Kocher et al., 1989).

### 2.2. DNA amplification and DNA sequencing

Fragments containing mtDNA CR and cyt *b* gene were obtained, respectively, by PCR amplifications. Primer

Table 1
Species and samples used in the present study and their GenBank accession numbers

| Classification | Species and haplotypes | Specimen voucher | Accession No. (Cyt *b*) | Accession No. (CR) |
|---|---|---|---|---|
| Botiinae | | | | |
| *Leptobotia* | *Leptobotia tchangi*1 | IHCAS0000024 | AY625719 | AY600871 |
| | *Leptobotia tchangi*2 | IHCAS0000025 | AY625720 | DQ105268 |
| | *Leptobotia tchangi*3 | IHCAS0000026 | AY625722 | DQ105269 |
| | *Leptobotia tientaiensis*1 | IHCAS0000027 | AY625725 | AY600865 |
| | *Leptobotia tientaiensis*2 | IHCAS0000028 | AY625724 | AY600866 |
| | *Leptobotia pellegrini*1 | IHCAS0000029 | AY625723 | AY600873 |
| | *Leptobotia pellegrini*2 | IHCAS0301046 | DQ105204 | DQ105270 |
| | *Leptobotia rubrilabris*1 | IHCAS0000021 | AY625716 | AY600872 |
| | *Leptobotia rubrilabris*2 | IHCAS0000022 | AY625717 | DQ105267 |
| | *Leptobotia elongata*1 | IHCAS0000023 | AY625714 | DQ105271 |
| | *Leptobotia elongata*2 | IHCAS 0000019 | AY625715 | AY600875 |
| | *Leptobotia taeniops* | IHCAS0000020 | AY625718 | AY600870 |
| | *Leptobotia hansuiensis* | IHCAS0307110 | DQ105205 | AY600874 |
| *Parabotia* | *Parabotia fasciata*1 | IHCAS0000032 | AY625709 | DQ105272 |
| | *Parabotia fasciata*2 | IHCAS0000038 | AY625710 | AY600868 |
| | *Parabotia banarescui* | IHCAS0000037 | AY625711 | AY600869 |
| | *Parabotia lijiangensis* | IHCAS0000036 | AY625713 | AY600867 |
| | *Parabotia kiangensis* | IHCAS0307108 | AY625712 | DQ105273 |
| *Botia* | *Botia supericiliaris*1 | IHCAS0000030 | AY625704 | AY600862 |
| | *Botia supericiliaris*2 | IHCAS0000031 | AY625702 | AY600863 |
| | *Botia supericiliaris*3 | IHCAS0307109 | AY625703 | DQ105274 |
| | *Botia robusta*1 | IHCAS0000033 | AY625707 | AY600864 |
| | *Botia robusta*2 | IHCAS0307114 | AY625708 | DQ105279 |
| | *Botia robusta*3 | IHCAS0301041 | DQ105208 | DQ105280 |
| | *Botia pulchra*1 | IHCAS0301007 | AY625705 | DQ105275 |
| | *Botia pulchra*2 | IHCAS0301008 | AY625706 | DQ105276 |
| | *Botia nigrolineata* | IHCAS0301045 | DQ105209 | DQ105281 |
| | *Botia* sp. 1 | IHCAS0301038 | DQ105206 | DQ105277 |
| | *Botia* sp. 2 | IHCAS0301039 | DQ105207 | DQ105278 |
| Cobitinae | | | | |
| | *Paramisgurnus dabryanus* | IHCAS0208007 | AY625701 | DQ105316 |
| | *Misgurnus bipartitus*1 | IHCAS0301016 | DQ105237 | DQ105309 |
| | *Misgurnus bipartitus*2 | IHCAS0301017 | DQ105239 | DQ105311 |
| | *Lepidocephalus octocirrhus* | IHCAS0000015 | DQ105245 | DQ105317 |
| | *Cobits macrostigma*1 | IHCAS0208004 | DQ105229 | DQ105301 |
| | *Cobits macrostigma*2 | IHCAS0307111 | DQ105230 | DQ105302 |
| | *Cobitis granoci* | IHCAS0301019 | DQ105242 | DQ105313 |
| | *Cobitis lutheri* | IHCAS0301021 | DQ105231 | DQ105303 |
| | *Misgurnus anguillicaudatus*1 | IHCAS0000003 | DQ105240 | AY600879 |
| | *Misgurnus anguillicaudatus*2 | IHCAS0000005 | DQ105241 | DQ105312 |
| | *Misgurnus anguillicaudatus*3 | IHCAS0000006 | DQ105238 | DQ105310 |
| | *Niwaella* cf. *laterimaculata* | IHCAS0000009 | DQ105236 | DQ105308 |
| | *Cobitis* cf. *sinensis*1 | IHCAS0000008 | DQ105234 | DQ105306 |
| | *Cobitis* cf. *sinensis*2 | IHCAS0000011 | DQ105233 | DQ105305 |
| | *Cobitis sinensis* | IHCAS0000012 | AY625699 | AY600880 |
| | *Cobitis* cf. *sinensis*3 | IHCAS0000013 | DQ105235 | DQ105307 |
| | *Cobitis* cf. *granoci* | IHCAS0000014 | DQ105243 | DQ105314 |
| | *Cobitis* cf. *taenia* | IHCAS0000017 | DQ105244 | DQ105315 |
| | *Cobitis* cf. *dolicorhynchus* | IHCAS0000018 | DQ105232 | DQ105304 |
| Nemacheilinae | | | | |
| | *Paracobitis variegatus* | IHCAS0301029 | AY625697 | DQ105265 |
| | *Paracobits potanini* | IHCAS0307106 | DQ105203 | DQ105266 |
| | *Barbatula nuda*1 | IHCAS0000043 | DQ105252 | DQ105324 |
| | *Barbatula nuda*2 | IHCAS0208022 | DQ105253 | DQ105325 |
| | *Barbatula barbatula*1 | IHCAS0307299 | DQ105254 | DQ105326 |
| | *Barbatula barbatula*2 | IHCAS0307181 | DQ105255 | DQ105327 |
| | *Triplophysa stenura*1 | IHCAS0000098 | DQ105247 | DQ105319 |
| | *Triplophyda stenura*2 | IHCAS0307104 | DQ105246 | DQ105318 |
| | *Triplophysa stewarti* | IHCAS0307103 | DQ105248 | DQ105320 |
| | *Triplophysa stoliczkae* | IHCAS0000099 | DQ105249 | DQ105321 |

(*continued on next page*)

Table 1 (*continued*)

| Classification | Species and haplotypes | Specimen voucher | Accession No. (Cyt *b*) | Accession No. (CR) |
|---|---|---|---|---|
| | *Triplophysa orientalis* | IHCAS0405365 | DQ105251 | DQ105323 |
| | *Nemacheilus subfuscus*1 | IHCAS0307101 | DQ105224 | DQ105296 |
| | *Nemacheilus subfuscus*2 | IHCAS0307102 | DQ105225 | DQ105297 |
| | *Nemacheilus putaoensis* | IHCAS0301002 | DQ105226 | DQ105298 |
| | *Nemacheilus polytaenia* | IHCAS0000045 | DQ105227 | DQ105299 |
| | *Micronemacheilus pulcher*1 | IHCAS0307112 | DQ105198 | DQ105259 |
| | *Micronemacheilus pulcher*2 | IHCAS0307113 | DQ105199 | DQ105260 |
| | *Lefura costata* | IHCAS0307107 | DQ105196 | DQ105257 |
| | *Triplophysa* sp. | IHCAS0307105 | DQ105250 | DQ105322 |
| | *Schistura thai* | IHCAS0000047 | DQ105202 | DQ105264 |
| | *Schistura fasciolata* | IHCAS0000049 | DQ105201 | DQ105263 |
| | *Schistura longa* | IHCAS0000050 | AY625698 | DQ105261 |
| | *Schistura kloetzliae* | IHCAS0000016 | DQ105228 | DQ105300 |
| | *Sectoria heterognathos* | IHCAS0301054 | DQ105200 | DQ105262 |
| | *Oreonectes platycephalus* | IHCAS0301039 | DQ105197 | DQ105258 |
| Balitorinae | | | | |
| | *Vanmanenia pingchowensis*1 | IHCAS0000064 | AY625727 | DQ105289 |
| | *Vanmanenia pingchowensis*2 | IHCAS0000066 | DQ105219 | DQ105290 |
| | *Crossostoma stigmata* | IHCAS0301049 | DQ105220 | DQ105291 |
| | *Beaufortia szechuanensis* | IHCAS0000096 | AY625726 | DQ105294 |
| | *Beaufortia kweichowensis* | IHCAS0301034 | DQ105223 | DQ105295 |
| | *Pseudogastromyzon tungpeiensis* | IHCAS0301047 | DQ105221 | DQ105292 |
| | *Pseudogastromyzon jiulongjiangensis* | IHCAS0301050 | DQ105222 | DQ105293 |
| | *Hemimyzon abbreviata* | IHCAS0307117 | DQ105211 | AY600876 |
| | *Hemimyzon sinensis* | IHCAS0307118 | DQ105210 | DQ105282 |
| | *Sinogastromyzon szechuanensis*1 | IHCAS0307119 | DQ105213 | AY600877 |
| | *Sinogastromyzon szechuanensis*2 | IHCAS0307120 | DQ105214 | DQ105285 |
| | *Siongastromyzon wui* | IHCAS0301040 | DQ105212 | DQ105284 |
| | *Sinogastromyzon hsiashiensis* | IHCAS0301052 | DQ105215 | DQ105286 |
| | *Lepturichthys fimbriata* | IHCAS0000088 | AY625695 | DQ105283 |
| | *Sinohomaloptera kwangsiensis* | IHCAS0307116 | DQ105216 | AY600878 |
| | *Balitora elongata*1 | IHCAS0301030 | DQ105217 | DQ105287 |
| | *Balitora elongata*2 | IHCAS0301053 | DQ105218 | DQ105288 |
| | *Metahomaloptera omeiensis* | IHCAS0000100 | DQ111990 | DQ112166 |
| Catostomidae | | | | |
| | *Myxocyprinus asiaticus* | | AF036176* | AY017140* |
| Gyrinocheilidae | | | | |
| | *Gyrinocheilus aymonieri* | IHCAS0301042 | DQ105256 | DQ105328 |
| Cyprinidae | | | | |
| | *Cyprinus carpio* | | NC001606* | NC001606* |
| | *Danio rerio* | | NC002333* | NC002333* |

An asterisk (*) denotes a sequence that was downloaded from GenBank. Nomenclature is according to Nelson (1994).

sets, DL1 (5′-ACC CCT GGC TCC CAA AGC-3′) and DH2 (5′-ATC TTA GCA TCT TCA GTG-3′) were designed for the CR (Liu et al., 2002), which is located in tRNA-pro and tRNA-phe, respectively. L14724 (5′-GAC TTG AAA AAC CAC CGT TG-3′) and H15915 (5′-CTC CGA TCT CCG GAT TAC AAG AC-3′) (Xiao et al., 2001) were used for cytochrome *b* gene. PCR was performed at an initial denaturation step at 94 °C for 3 min, followed by 35 cycles at 94 °C for 30 s, 52–58 °C for 45 s, 72 °C for 1 min, and a final extension at 72 °C for 8 min. The amplified fragments were purified with BioStar glassmilk DNA purification kit following the manufacture's instruction. The purified fragments were sequenced by Shanghai DNA Biotechnologies company. All sequences are available from GenBank (accession numbers are listed in Table 1).

### 2.3. Sequence analysis

Nucleotide sequences were aligned using Clustal X (Thompson et al., 1997) and refined manually with SEAVIEW (Galtier et al., 1996). Base compositional bias and sequence divergences were calculated and a chi-square ($\chi^2$) test of base heterogeneity was conducted using PAUP* version 4.0b10 (Swofford, 2002) for all positions. Nucleotide saturation was analyzed by plotting absolute number of transitions ($T_i$) and transversions ($T_v$) against HKY distance values in PAUP*.

Combined data were analyzed by maximum parsimony (MP), neighbor-joining (NJ), and Bayesian methods for phylogenetic reconstruction. Congruence among tree topologies generated with cyt *b* and CR sequences was tested with the incongruence length difference test (ILD) as

implemented in the partition homogeneity test in PAUP* (Farris et al., 1994; Mickevich and Farris, 1981). Modeltest 3.06 (Posada and Crandall, 1998) was used to determine the best-fit evolutionary model for NJ and Bayesian analysis, and a hierarchical series of likelihood ratio tests (LRTs) was performed using this program. MP and NJ analyses were conducted using PAUP*. Bayesian analysis was carried out using MrBayes version 3.0b (Huelsenbeck and Ronquist, 2001). A heuristic search was used to estimate the most likely topology for NJ and MP methodologies. Heuristic searches started with stepwise addition tree; branch swapping was performed by the tree-bisection–reconnection (TBR) method using default parameters. Bootstrap analysis with 1000 replications was used to estimate support for the resulting topologies.

In Bayesian analysis, starting trees were random. Four simultaneous Markov chains were run for 1,000,000 generations. Trees were sampled after every 100 generations, with a total of 10,001 trees. Stationarity was read after 100,000 generations. Therefore, the first 1000 trees were ignored and the posterior probability of the phylogeny was determined from the resulting 9001 trees. Two independent Bayesian analyses were performed to check for local optima.

## 3. Results

### 3.1. Base compositions

Following alignment of the 1140 bp of cyt *b* gene obtained for 95 individuals (including outgroups), no deletions or insertions were observed. Plots of the number of substitutions against HKY distances revealed no saturation for $T_i$ or $T_v$ for all positions (not shown). Base frequencies were heterogenous across all taxa for all three codon positions ($\chi^2 = 383.168$, df = 282, $P = 0.000 < 0.001$). Nucleotide composition at the third position exhibited significant heterogeneity: first position, $\chi^2 = 43.789$, df = 282, $P = 1.000$; second position, $\chi^2 = 5.598$, df = 282, $P = 1.000$; and third position, $\chi^2 = 1305.826$, df = 282, $P = 0.000 < 0.001$. Among the 1140 bp, 562 sites were variable, of which 513 were parsimony informative. The average nucleotide composition for all the sequences was A = 28.1%, T = 28.7%, C = 28.2%, and G = 15.0%. The content of A + T (56.8%) was higher than that of C + G (43.2%). Strong compositional biases against G existed at the third position (only 5.5%). $T_i$ outnumbered $T_v$ at all levels of sequence divergence, and the average $T_i/T_v$ ratio was 2.093.

For CR, the length in our sampled specimens ranged from 834 to 944 bp and many indels were observed. There was no significant difference in base frequencies across all taxa ($\chi^2 = 143.534$, df = 282, $P = 1.000$). Plots of the number of substitutions against HKY distances showed that both $T_i$ and $T_v$ had not reached saturation (not shown). The average base composition was A = 34.5%, T = 31.9%, C = 19.8%, and G = 13.8%. Compared to the cyt *b* gene, CR showed a strong bias in base content with two times higher content of A + T (66.4%) than C + G (33.6%). As in other fishes (Zhu et al., 1994), $T_i$ outnumbered $T_v$ in comparisons between closely related samples, but between the more divergent sequences, $T_v$ was equal to or more than $T_i$. The average $T_i/T_v$ ratio was 1.001.

### 3.2. Comparison of evolutionary rates in the cyt b and CR

The statistical analysis of sequence divergences for 95 individuals indicated that the HKY distance for the cyt *b* was 0.000–32.3 and 1.1–67.0% for the CR (data not shown). Divergences among the CR sequences were broadly higher than those of the cyt *b*. A graphic comparison of pairwise corrected sequence divergences (HKY distance) for the cyt *b* and CR was shown in Fig. 2, which revealed a linear relation between the two segments and indicated that generally the CR sequence is diverging faster than the cyt *b* gene sequence (the ratio of CR/cyt *b* is 1.83) for the same set of taxa. However, considering only the more closely related sequences, i.e., those within the cyt *b* divergence of <10% and close phylogenetic relatives based on analysis, the CR segment is evolving slower than cyt *b* gene (CR/cyt *b* = 0.78).

Considering the relationship between sequence variations and current classification, the levels of sequence divergence are closely related to the rank of the existing classification in the Cobitoidea. Fishes from different populations of the same species have divergence of <6.90% in cyt *b* and <4.80% in CR for most sequences. Unusual among these comparisons across taxa is the divergence between *Hemimyzon abbreviata*, *Hemimyzon sinensis*, and *Lepturichthys fimbriata*, wherein divergence in cyt *b* is minimal, ranging from 2.0 to 2.9%, despite the fact that those species have marked differences in morphological characters. Species of different genera also display notable difference in levels of divergence, such as >7.79% in cyt *b* and >5.16% in CR between *Parabotia* and *Leptobotia*, while difference between *Parabotia* and *Botia* are >14.80% in cyt *b* and >24.64% in CR. Among the subfamilies, divergence in CR is constantly larger than that in cyt *b*.

### 3.3. Phylogenetic analysis

A total of 2174 bp (including gaps in the CR segment) were analyzed for each of 95 individuals (including outgroups). Among 2174 bp, 1410 bp were variable and 1254 bp were parsimony informative. Base frequencies were homogeneous across all sites and did not differ significantly among all specimens ($\chi^2 = 329.622$, df = 276, $P = 0.015$). The partition homogeneity test revealed no significant differences among any of the segments studied (cyt *b* versus CR, $P = 0.07 > 0.01$; Cunningham, 1997). Plots of the absolute numbers of transitions and transversions against HKY distance revealed no trend towards some level of saturation (not shown). The average $T_i/T_v$ ratio was 1.331.

All inference methods yielded very similar topologies of combined sequence data (Figs. 3–5) with a few variations

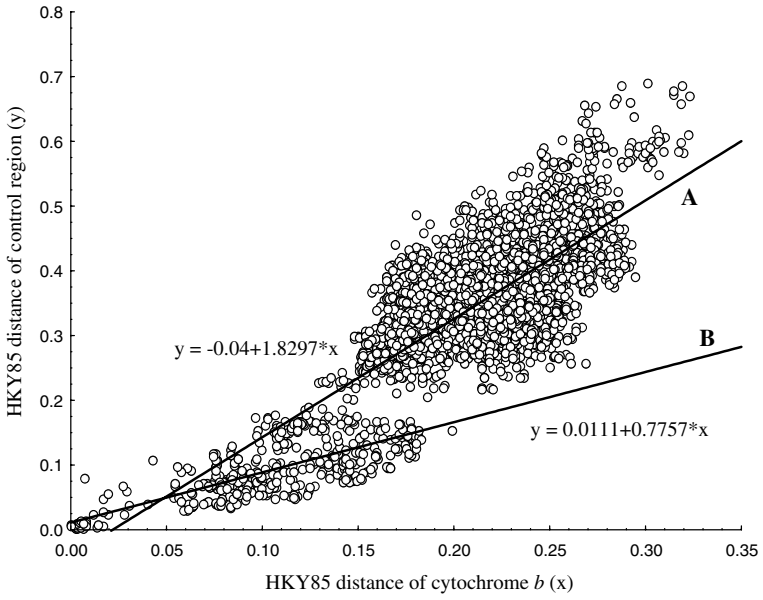


Fig. 2. HKY distance of cytochrome *b* vs. control region. Line A represents the relationship of overall sequence divergences between two segments and indicates that generally the CR sequence is diverging faster than the cyt *b* gene sequence (CR/cyt *b* = 1.83). Line B just includes the points for which the cyt *b* sequence divergence is 0–10% and indicates that the CR segment is evolving slower than cyt *b* gene (CR/cyt *b* = 0.78).

occurring at basal nodes and between a few species. MP analysis employed an equal weighting scheme of $T_v$ and $T_i$ and all positions were included. Bootstrap consensus in two equally parsimonious trees (Fig. 3) was obtained with a tree length of 11,777 steps, CI = 0.230, RI = 0.671. The MP tree indicated that the phylogenetic relationship of the Cobitoidea was ((Catostomidae + Gyrinocheilidae) + (Botiinae + (Balitorinae + (Cobitinae + Nemacheilinae)))), which suggested that Sawada's Cobitidae (including Cobitinae and Botiinae) was not monophyletic and his Balitoridae (including Nemacheilinae and Balitorinae) did not cluster together. However, these four subfamilies each formed their own monophyletic groups, respectively, with high bootstrap values. Nemacheilinae and Cobitinae formed a clade that was sister to Balitorinae, and these two clades formed a large lineage that was sister to the basal-most lineage Botiinae. Within the Botiinae, three independent groups were included, representing the three genera, *Leptobotia*, *Parabotia*, and *Botia*. Each genus was resolved as a monophyletic group with corresponding bootstrap values of 73, 68, and 100. *Leptobotia* and *Parabotia* were sister taxa and this clade was sister to *Botia*. The relationships among the species of the Cobitinae are complicated, the genus *Misgurnus* is nested within *Cobitis*, which was divided into two groups. The clade Balitorinae was divided into two clades, corresponding to Hora's Gastromyzoninae and Homalopterinae (Hora, 1932). Within Nemacheilinae, our analysis included limited samples of genera; however, all of the genera sampled were resolved as monophyletic and most of the generic species relationship was highly supported.

Based on Modeltest, the HKY model with an estimate of invariable sites (0.308) and a discrete approximation of the gamma distribution (0.947) was chosen. Using this model, we obtained one NJ tree with NJ analysis (Fig. 4). The NJ tree indicated that the phylogenetic relationship of the Cobitoidea was (((Catostomidae + Gyrinocheilidae) + Botiinae) + (Balitorinae + (Cobitinae + Nemacheilinae))), which suggested that neither the Cobitidae (including Cobitinae and Botiinae) nor the Balitoridae (including Nemacheilinae and Balitorinae) formed monophyletic group. As in the MP analysis, the four subfamilies formed their own monophyletic group, respectively, with high bootstrap values (78 in Botiinae, 95 in Balitorinae, 100 in Cobitinae, and 99 in Nemacheilinae). Different from the MP tree, the clade (Catostomidae + Gyrinocheilidae) clustered with the Botiinae, and formed a larger clade that was sister to the other loaches. For each subfamily, the topology of NJ tree is almost congruent with MP tree except for the branching order of a few samples.

Two independent Bayesian analyses produced the same topology with slight differences in posterior probabilities. Herein, we provide one of these trees (Fig. 5). As with MP and NJ analyses, the monophyly of the four subfamilies was recovered and was supported with high posterior probabilities (1.00 in every subfamily). The topology within each subfamily is similar to MP and NJ analyses. Botiinae is distantly related to the other three subfamilies in the Cobitoidea, which is also supported by MP and NJ tree. The phylogenetic relationships among four subfamilies are the same as in the MP analyses.

## 4. Discussion

### 4.1. Dynamics of the evolutionary rate of CR

Generally, the CR sequences evolve more rapidly than cyt *b* sequences in the Cobitoidea, however, in considering

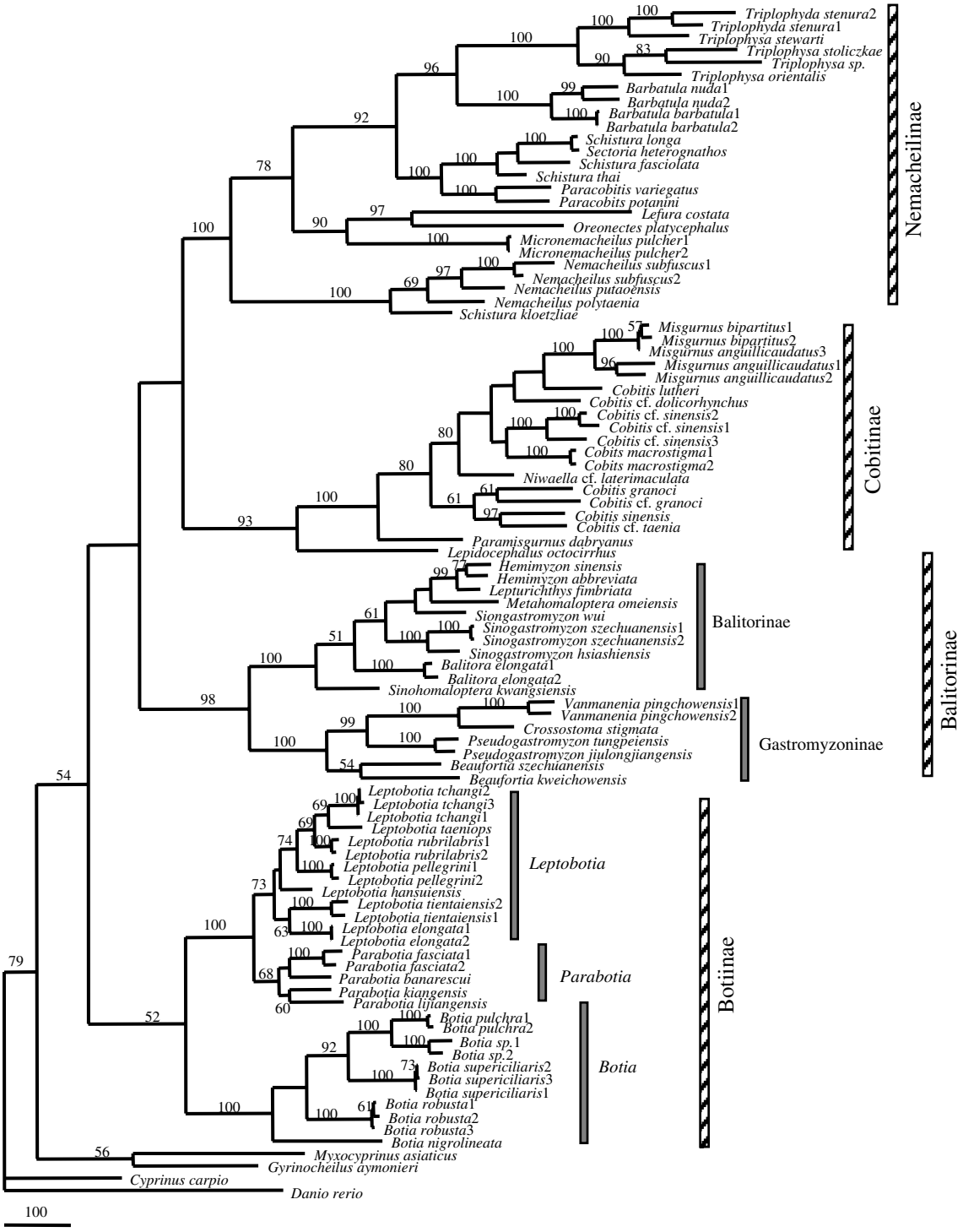


Fig. 3. Phylogeny of the Cobitoidea based on maximum parsimony (MP) analysis of combined cytochrome *b* and control region sequences. Numbers above the nodes represent bootstrap values with 1000 replications. Only values $\geqslant 50$ are reported.

only closely related species, CR sequence evolution was slower than that of cyt *b* gene. Roukonen and Kvist (2002) reported a similar finding in a survey of 68 avian species. They proposed that the trend of the ratio of CR versus cyt *b* divergences seems to be somewhat genus specific; many avian lineages were shown to have more rapidly evolving CR (e.g., among the *Cyanoramphus* species, 5.14–21.65 times faster), but within the genus *Alectoris* and *Polioptila*, CR/cyt *b* ratios were less than 1 (0.46:0.94 and 0.36:0.81, respectively).

Saunders and Edwards (2000) studied dynamics and phylogenetic implications of mtDNA CR sequences in the New World Jays and found a slow rate of evolution in the CR. They suggested that their data indicated a higher

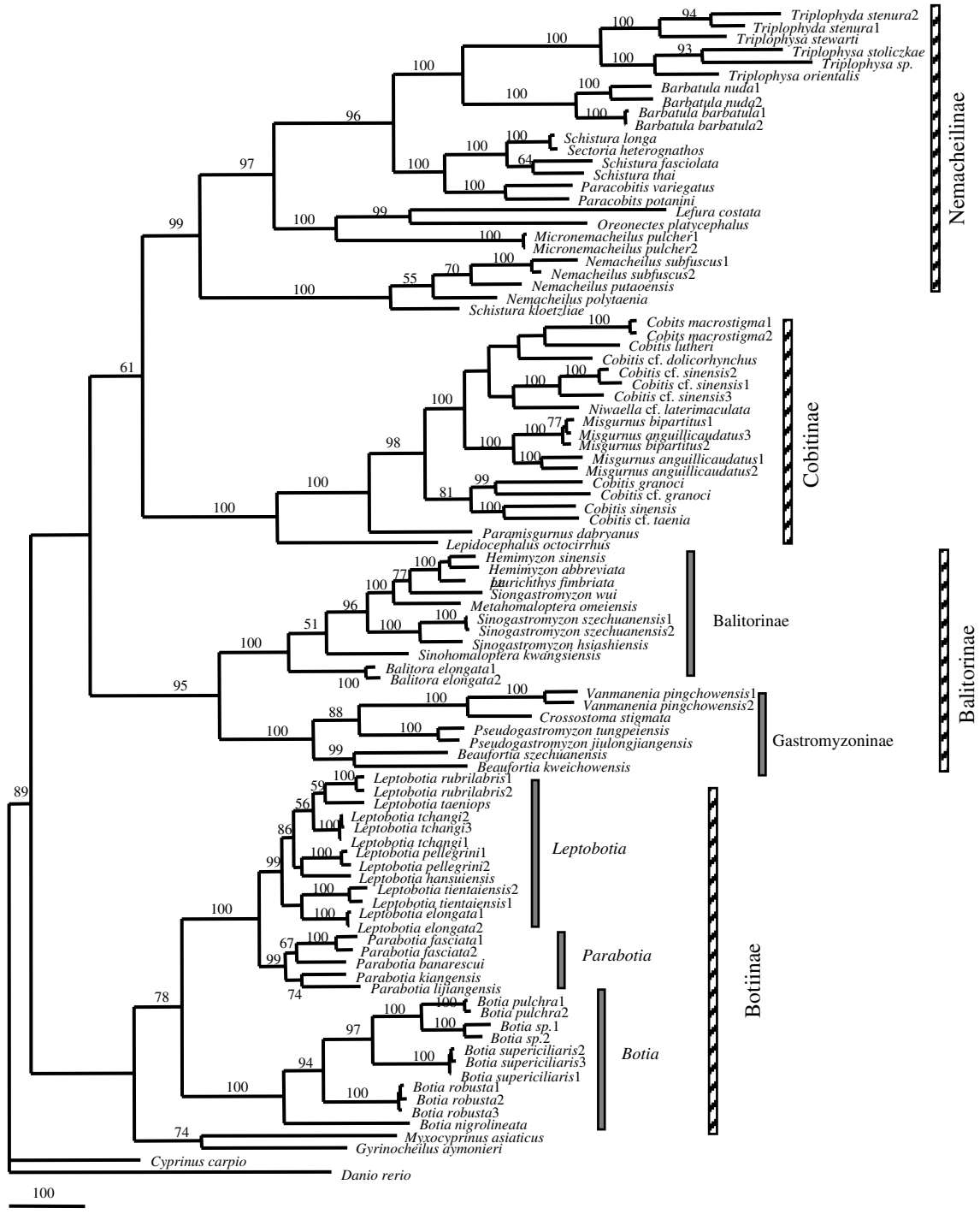


Fig. 4. Phylogeny of the Cobitoidea based on neighbor-joining (NJ) analysis of combined cytochrome *b* and control region sequences. Numbers above the nodes represent bootstrap values with 1000 replications. Only values $\geqslant 50$ are reported.

level of selective constraint in control domain I than in the third positions of cyt *b*. Studies have showed that CR contains sequences related to termination of H-strand replication, the origin of H-strand, and promoters of transcription to both L- and H-strand (Doda et al., 1981; Randi and Lucchini, 1998; Saccone et al., 1991; Sbisà et al., 1997; Southern et al., 1988). This indicates that the CR has evolutionary constraints. Besides, many conserved sequence blocks identified suggest that many unknown functions exist. It is these known and unknown functions that put the CR under high evolutionary pressure and lead to the slow rate of substitution. The ability to fold into secondary structures is essential for function of the origin of replication of many systems and the termination of transcription of RNA (Brown et al., 1986). Because of its function, it is easy to understand that the CR contains sequences that can fold

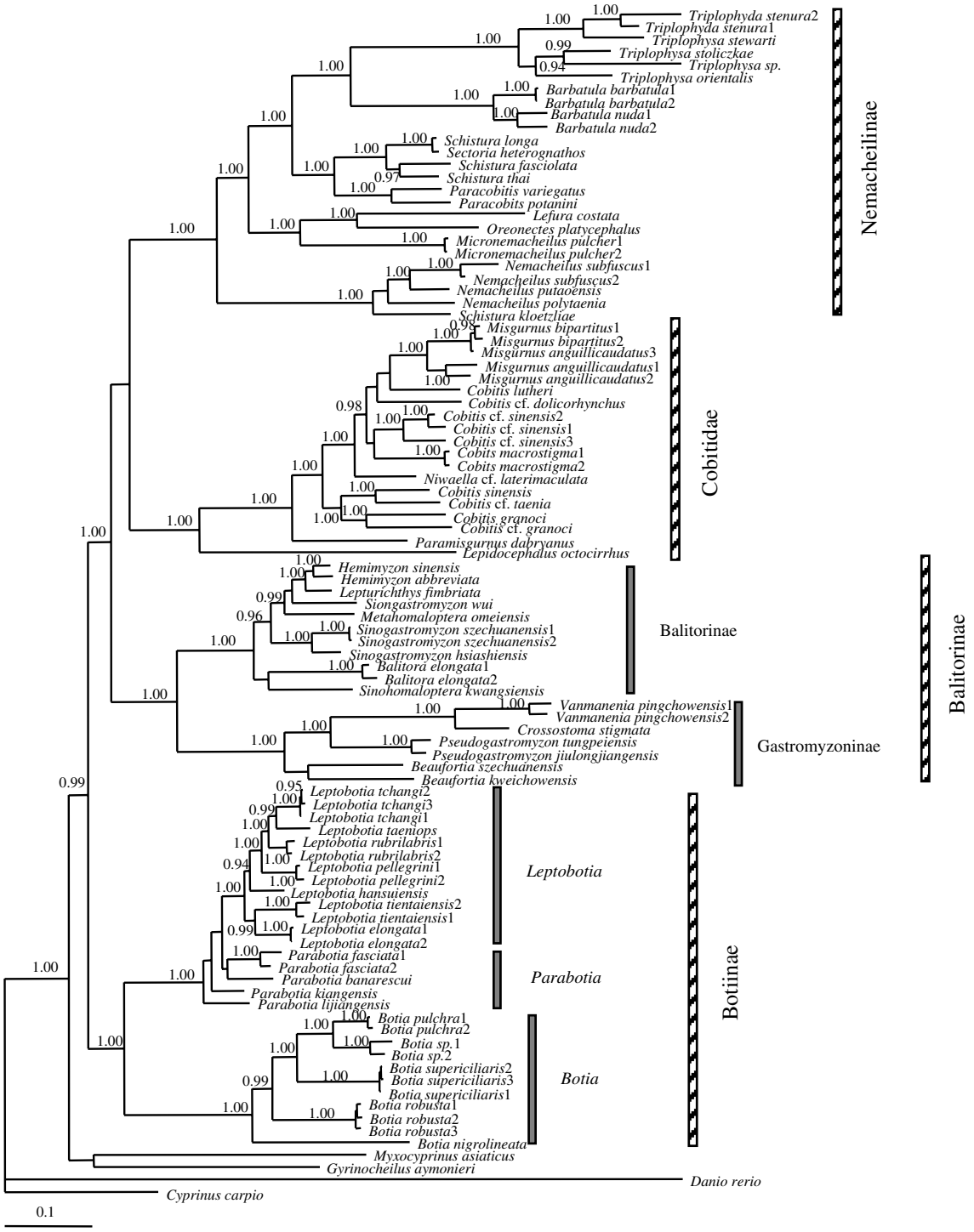


Fig. 5. Phylogeny of the Cobitoidea based on 50% major rule consensus tree obtained from Bayesian analysis of combined cytochrome *b* and control region sequences. Numbers above the nodes are Bayesian posterior probabilities. Only values $\geqslant 0.80$ are reported.

into secondary structure. Folding into a secondary structure can help preserve the functionality of the sequences, however, this is not the primary reason for slow rate of the CR divergence. Many functions and evolutionary constraints are likely the main reasons for this conservation. Many studies have demonstrated that the CR in vertebrates shows similar structure and conserved sequences (Lee et al., 1995; Randi and Lucchini, 1998; Sbisà et al., 1997; Southern et al., 1988), indicating evolutionary constraints and conservatism at various levels.

### 4.2. Phylogenetic implications of the mtDNA cyt b gene and CR

Rychel et al. (2004) mentioned that a better estimate of the true phylogeny may be obtained and/or overall clade

support may be improved by combining data into a single analysis. However, it is still a contentious issue as to whether data can be or should be combined. Bull et al. (1993) were against combining data partitions if heterogeneity is known to exist between them, while Wiens (1998) demonstrates that localized areas of conflict between data sets may not disrupt overall analyses, and in areas of data congruence, combining data strengthens the overall accuracy of the analysis. In our study, the partition homogeneity test between the cyt *b* and CR revealed no significant differences ($P = 0.07 > 0.01$), when using an adjusted $\alpha$ of 0.01 as suggested by Cunningham (1997). Phylogenetic analyses using the combined data also resulted in better topology structure than individual gene sequences, just as Rychel et al. (2004) indicated.

The topologies recovered by analysis of combined data using the three methods herein reject the hypothesis of Sawada (1982) who suggested that the Cobitidae and Balitoridae evolved separately as a monophyletic group. MP and BI trees supported the Botiinae as the basal-most clade for the loaches, what is consistent with the conclusion of Liu et al. (2002). Regardless of the positions of Catostomidae and Gyrinocheilidae, three analyses resolved well-supported monophyletic subfamilies.

### 4.3. Systematic implications in the superfamily Cobitoidea

Traditionally, Cobitinae, Botiinae, Nemacheilinae, and Balitorinae were recognized as subfamilies included in the Cobitoidea. Nalbant and Bianco (1998) indicated that in the study of Sawada (1982) many osteological similarities between Balitorinae and Nemacheiliane are due to the homoplasies, so they proposed that Nemacheilinae should be considered a distinct family, the Nemacheilidae, which together with the families Cobitidae and Botiidae, is included in the superfamily Cobitoidea. The Balitoridae is also regarded as a distinct family. Our molecular data agreed with this opinion.

As seen from the topologies yielded by our data, there is no doubt that the Botiinae can be elevated to the family Botiidae. In another contribution (Tang et al., 2005), we have discussed this conclusion in more detail. To balance the rank of the taxonomy within the Cobitoidea, the other three subfamilies are also elevated to families.

The Cobitidae (sensu Nalbant) probably is, in present acceptance, a monophyletic group, which is consistent with osteological analyses by Sawada (1982). However, intergeneric and congeneric phylogenetic relationships are complex, especially for the genus *Cobitis*.

The Nemacheilidae is the largest group in the Cobitoidae, including numerous morphologically similar species and many taxonomic problems remain at the species level. Nalbant and Bianco (1998) thought that this clade probably had a polyphyletic origin, a conclusion not supported herein. However, some genera within the family are polyphyletic, such as *Schistura*, which includes several rather distinct groups of species that are difficult to delimit (Bănărescu and Nalbant, 1995). Our molecular phylogenetic trees show that all Nemacheilidae fishes clustered together. The phylogenetic relationships of the Nemacheilidae are in need of further analysis, with as many samples as possible.

As for the Balitoridae, when Hora (1932) first defined the group, he concluded that the family was polyphyletic in origin, with members of the Balitorinae having evolved from the Cyprinidae and members of Gastromyzoninae evolved from the Cobitidae (sensu Regan). Phylogenetic analyses in the current study refute this hypothesis as these two subfamilies form a monophyletic group. Thus, the Balitoridae is a monophyletic group, likely derived from the ancestor of the Nemacheilidae and Cobitidae, and it is divided into two subfamilies, Gastromyzoninae and Balitorinae, corresponding to Hora's Gastromyzoninae and Homalopterinae (Hora, 1932).

So, with the change of the systematic position of loach subgroups, based on our data, we suggest that the classification of the Cypriniformes is changed as following:

Cypriniformes
- Cyprinoidea
  - Cyprinidae
- Cobitoidea
  - Catostomidae
  - Gyrinocheilidae
  - Botiidae
  - Balitoridae
    - Gastromyzoninae
    - Balitorinae
  - Cobitidae
  - Nemacheilidae

## Acknowledgments

We are grateful to L. Zhang, J. Yang, H. Hu, Z. Peng, X. Guo, and L. Yang for help in collecting specimens. Thanks are also to Dr. M. Kottelat for advices on species identification and to the anonymous reviewers for their suggestions to this paper. Funding support was provided by Grants of NSFC 30170137, State Key project of NSFC 40432003, and Innovation Project of the Chinese Academy of Sciences (Grant No. KZCX-SW-126). Our work is also part of the project, the Cypriniformes Tree Of Life, funded by NSF of the USA.

## References

Bănărescu, P.M., Nalbant, T.T., 1995. A generical classification of Nemacheilinae with description of two new genera (Teleostei: Cypriniformes: Cobitidae). Trav. Mus. Hist. Nat. "Grigore Antipa" 35, 429–496.

Berg, L.S., 1940. Classification of fishes, both recent and fossil. Trav. Inst. Zool. Acad. Sci. USSR 5, 87–345.

Bernatchez, L., Danzmann, R.G., 1993. Congruence in control-region sequence and restriction-site variation in mitochondrial DNA of brook charr (Salvelinus fontinalis Mitchill). Mol. Biol. Evol. 10, 1002–1014.

Brown, W., 1985. The mitochondrial genome of animals. In: Macinttyre, R. (Ed.), Molecular Evolutionary Genetics. Plenum Press, New York, pp. 95–130.

Brown, G.G., Gadaleta, G., Pepe, G., Saccone, C., Sbisa, E., 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J. Mol. Biol. 192, 503–511.

Bull, J.J., Huelsenbeck, J.P., Cunning, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combing data in phylogenetic analysis. Syst. Biol. 42, 384–397.

Chen, J., Zhu, S., 1984. Phylogenetic relationships of subfamilies in the loach family Cobitidae (Pisces). Acta Zootaxon. Sin. 9, 201–207.

Crochet, P.A., Desmarais, E., 2000. Slow rate of evolution in the mitochondiral control region of Gulls (Aves: Laridae). Mol. Biol. Evol. 17, 1797–1806.

Cunningham, C.W., 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. 14, 733–740.

Doda, J.N., Wright, C.T., Clayton, D.A., 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA 78, 6116–6120.

Farris, J.S., Källerjö, M., Kluge, A.G., Butt, C., 1994. Testing significance of incongruence. Cladistics 10, 315–319.

Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO-WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12, 543–548.

He, S.P., Yue, P.Q., Chen, Y.Y., 1997. Comparative study in the morphology and development of the pharyngeal dentition in the families of Cypriniformes. Acta Zootaxon. Sin. 43, 255–262.

Hora, S.L., 1932. Classification, bionomics and evolution of Homalopteridae fishes. Mem. Indian Mus. 12, 263–330.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

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

Kottelat, M., 1988. Indian and Indochinese species of Balitora (Osteichthyes: Cypriniformes) with description of two new species and comments on the family-group names Balitoridae and Homalopteridae. Rev. Suisse. Zool. 95, 487–504.

Kottelat, M., 2001. Fishes of Laos. WHT, Colombo, Sri Lanka.

Lee, W., Conroy, J., Howell, W.H., Kocher, T.D., 1995. Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. 41, 54–66.

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.

Liu, H., Tzeng, C.S., Teng, H.Y., 2002. Sequence variations in the mitochondrial DNA control region and their implications for the phylogeny of the Cypriniformes. Can. J. Zool. 80, 569–581.

Meyer, A., 1993. Evolution of mitochondrial DNA in fishes. In: Hochachka, P.W., Mommsen, T.P. (Eds.), Biochemistry and Molecular Biology of Fishes, Vol. 2. Elsevier, Amsterdam, pp. 1–38.

Mickevich, M.F., Farris, J.S., 1981. The implications of congruence in menidia. Syst. Biol. 30, 351–370.

Moum, T., Árnason, E., 2001. Genetic diversity and population history of two related seabird species based on mitochondrial DNA control region sequences. Mol. Ecol. 10, 2463–2478.

Nalbant, T., 1963. A study of the genera of Botiinae and Cobitinae (Pisces, Ostariophysi, Cobitidae). Trav. Mus. Nat. "Grigore Antipa" 4, 343–375.

Nalbant, T., 2002. Sixty million years of evolution. Part one: family Botiidae (Pisces: Ostariophysi: Cobitidae). Trav. Mus. Nat. "Grigore Antipa" 44, 343–379.

Nalbant, T.T., Bianco, P.G., 1998. The loaches of Iran and adjacent region with description of six new species (Cobitoidea). Ital. J. Zool. 65, 109–123.

Nelson, J.S., 1994. Fishes of the World, third ed. Wiley, New York.

Peng, Z., He, S., Zhang, Y., 2004. Phylogenetic relationships of glyptosternoid fishes (Siluriformes: Sisoridae) inferred from mitochondrial cytochrome *b* gene sequence. Mol. Phylogenet. Evol. 31, 979–987.

Perdices, A., Cunha, C., Coelho, M.M., 2004. Phylogenetic structure of *Zacco platypus* (Teleostei, Cyprinidae) populations on the upper and middle Chang Jiang ( = Yangtze) drainage inferred from cytochrome *b* sequences. Mol. Phylogenet. Evol. 31, 192–203.

Posada, D., Crandall, K.A., 1998. Model test: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Ramaswami, L.S., 1953. Skeleton of cyprinoid fishes in relation to phylogenetic studies. V. The skull and the gasbladder capsule of the Cobitidae. Proc. Natl. Inst. Sci. India 18, 519–538.

Randi, E., Lucchini, V., 1998. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. J. Mol. Evol. 47, 149–162.

Regan, C.T., 1911. The classification of the teleostean fishes of the order Ostariophysi I. Cyprinidae. Ann. Mag. Nat. Hist. Ser. 8, 13–32.

Roukonen, M., Kvist, L., 2002. Structure and evolution of the avian mitochondrial control region. Mol. Phylogenet. Evol. 23, 422–432.

Rychel, A.L., Reeder, T.W., Berta, A., 2004. Phylogeny of mysticete whales based on mitochondrial and nuclear data. Mol. Phylogenet. Evol. 32, 892–901.

Saccone, C., Pesole, G., Sbisà, E., 1991. The main regulatory region of mammalian mitochondrial DNA: structure–function model and evolutionary pattern. J. Mol. Evol. 33, 83–91.

Saunders, M.A., Edwards, S.V., 2000. Dynamics and phylogenetic implications of mtDNA control region sequences in New World Jays (Aves: Corvidae). J. Mol. Evol. 51, 97–109.

Sawada, Y., 1982. Phylogeny and zoogeography of the superfamily Cobitoidea (Cyprinoidei, Cypriniformes). Memoirs of the Faculty of Fisheries, Hokkaido University 28, 65–223.

Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G., Sacone, C., 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implifications. Gene 205, 125–140.

Shedlock, A.M., Parker, J.D., Crispin, D.A., Pietsch, T.W., Burmer, G.C., 1992. Evolution of the salmonid mitochondrial control region. Mol. Phylogenet. Evol. 1, 179–192.

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

Southern, S.O., Southern, P.J., Dizon, A.E., 1988. Molecular characterization of a cloned dolphin mitochondrial genome. J. Mol. Evol. 28, 32–42.

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (* and other Methods), Version 4. Sinauer, Sunderland, MA.

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.

Taylor, M.F., McKechnie, S.W., Pierce, N., Kreitman, M., 1993. The Lepidopteran mitochondrial control region: structure and evolution. Mol. Biol. Evol. 10, 1259–1272.

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

Wiens, J.J., 1998. Combing data sets with different phylogenetic histories. Syst. Biol. 47, 568–581.

Wu, X., Chen, Y., Chen, X., Chen, J., 1981. A taxonomical system and phylogenetic relationship of the families of the suborder Cyprinoidei (Pisces). Sci. Sin. 24, 563–572.

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.

Zhu, D., Jamieson, B.G., Hugall, A., Moritz, C., 1994. Sequence evolution and phylogenetic signal in control-region and cytochrome *b* sequences of rainbow fishes (Melanotaeniidae). Mol. Biol. Evol. 11, 672–683.