Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 47 (2008) 472–487
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Variation patterns of the mitochondrial 16S rRNA gene with secondary structure constraints and their application to phylogeny of cyprinine fishes (Teleostei: Cypriniformes)

Junbing Li [a,b], Xuzhen Wang [a], Xianghui Kong [c], Kai Zhao [d], Shunping He [a,*], Richard L. Mayden [e,1]

[a] *Laboratory of Fish Phylogenetics and Biogeography, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, People's Republic of China*
[b] *Graduate School of Chinese Academy of Sciences, Beijing 100039, People's Republic of China*
[c] *College of Life Science, Henan Normal University, Xinxiang 453007, China*
[d] *Agriculture and Animal Husbandry College, Qinghai University, Xining 810003, China*
[e] *Department of Biology, 3507 Laclede Avenue, Saint Louis University, St. Louis, MO 63103, USA*

Received 7 March 2007; revised 9 September 2007; accepted 14 September 2007
Available online 21 September 2007

Abstract

The mitochondrial 16S ribosomal RNA (rRNA) gene sequences from 93 cyprinid fishes were examined to reconstruct the phylogenetic relationships within the diverse and economically important subfamily Cyprininae. Within the subfamily a biased nucleotide composition (A > T, C > G) was observed in the loop regions of the gene, and in stem regions apparent selective pressures of base pairing showed a bias in favor of G over C and T over A. The bias may be associated with transition–transversion bias. Rates of nucleotide substitution were lower in stems than in loops. Analysis of compensatory substitutions across these taxa demonstrates 68% covariation in the gene and a logical weighting factor to account for dependence in mutations for phylogenetic inference should be 0.66. Comparisons of varied stem–loop weighting schemes indicate that the down-weightings for stem regions could improve the phylogenetic analysis and the degree of non-independence of stem substitutions was not as important as expected. Bayesian inference under four models of nucleotide substitution indicated that likelihood-based phylogenetic analyses were more effective in improving the phylogenetic performance than was weighted parsimony analysis. In Bayesian analyses, the resolution of phylogenies under the 16-state models for paired regions, incorporating GTR + G + I models for unpaired regions was better than those under other models. The subfamily Cyprininae was resolved as a monophyletic group, as well as tribe Labein and several genera. However, the monophyly of the currently recognized tribes, such as Schizothoracin, Barbin, *Cyprinion* + *Onychostoma* lineages, and some genera was rejected. Furthermore, comparisons of the parsimony and Bayesian analyses and results of variable length bootstrap analysis indicates that the mitochondrial 16S rRNA gene should contain important character variation to recover well-supported phylogeny of cyprinid taxa whose divergences occurred within the recent 8 MY, but could not provide resolution power for deep phylogenies spanning 10–19 MYA.

Keywords: Mitochondrial 16S rRNA gene; Secondary structure constraints; Phylogeny; Cyprininae

1. Introduction

The subfamily Cyprininae (sensu Howes, 1991) is one of the most diverse groups in the family Cyprinidae (Howes, 1991; Banarescu and Coad, 1991), with approximately

* Corresponding author. Fax: +86 27 68780071.
E-mail addresses: clad@ihb.ac.cn (S. He), cypriniformes@gmail.com (R.L. Mayden).
[1] Fax: +1 314 977 3658.

1055-7903/$ - see front matter © 2008 Published by Elsevier Inc.
doi:10.1016/j.ympev.2007.09.012

1332 species (Nelson, 2006). As an Old World clade, this group occurs from Europe and Africa to India, East Asia, and South East Asia (including most of Indonesia) (Chen et al., 1984; Howes, 1991). Cyprinine fishes are also among the largest representatives of the Order Cypriniformes, ranging in total length from about 3.1 to 120 cm as adults, and occupy habitats ranging from warm tropical waters to cold-water northern or mountainous streams and rivers. This group includes some of the most economically important freshwater fishes worldwide in terms of aquaculture and their impacts on ecosystems as exotic species. Many of these species serve as fundamental sources of protein in some cultures (60–80% of protein resource) (species of *Barbus*, *Catla*, *Cirrhinus*, *Ctenopharyngodon*, *Cyprinus*, *Hypothalmichthys*, and *Labeo*). Alternatively, many of these same species of been transplanted to continents and drainages where they are not native and their presence in these new ecosystems has impacted the native floras and faunas. The tremendous diversity of species, morphologies, and ecologies in this lineage makes it particularly important for research aimed at understanding various evolutionary, ecological, and biogeographical hypotheses and theories, mechanisms and rates of anagenesis and speciation, and testing alternative phylogenetic hypotheses of the Cyprinidae. However, the evolutionary history of the Cyprininae is poorly known expect for morphological studies corroborating the monophyly of the subfamily and some small groupings within (Gosline, 1978; Chen et al., 1984; Howes, 1991; Rainboth, 1991; Cavender and Coburn, 1992).

The systematic history of cyprinine fishes has been marked with traditional morphological studies but cursory examination of the total diversity of the group that has resulted in varied classifications. Based on external morphology and other anatomical characters, Chen et al. (1984) recognized the Tribe Barbini (= Cyprininae sensu Howes, 1991) as including the Labeoninae, Barbinae, and Cyprininae. Rainboth (1991) classified Cyprininae into many tribes and subtribes. Considering the distribution of barbels, morphotype, and innervation of barbels, Howes (1991) recognized six lineages in his Cyprininae, the Barbins, Labeins, Squaliobarbins, Schizothoractins, *Cyprinion–Onychostoma*, and other taxa (containing *Barbus sensu lato*, *Puntius*, *Tor*, *Oreodaimon*, *Pseudobarbus*, and *Gibelion*). Chen (1998) subdivided cyprinine fishes into four subfamilies, including Barbinae, Labeoninae, Cyprininae, and Schizothoracinae.

Despite varied efforts to understand the phylogenetic relationships of the family Cyprinidae using molecular methods (Durand et al., 2002; Briolay et al., 1998; Gilles et al., 1998, 2001; Zardoya and Doadrio, 1998, 1999; Wang et al., 2002, 2007; He et al., 2004a,b; Liu and Chen, 2003; Li et al., 2005), relationships among most genera or tribes of Cyprininae remain unresolved or are inconsistent between existing phylogenies. Molecular phylogenetic investigations of the subfamily have focused only on subgroups, including species-level analyses (Tsigenopoulos et al., 1999; Tsigenopoulos and Berrebi, 2000; He et al., 2004a,b), genus-level analyses (He et al., 2004a,b; Li et al., 2005), and intertribal evaluations (Cunha et al., 2002; Liu and Chen, 2003; He et al., 2004a,b; Wang et al., 2007). All of these analyses have failed to corroborate lineages previously identified in morphological-based classifications or resolve deep-level relationships in the subfamily. The poor resolution of relationships has largely been due to limited taxonomic sampling and/or limited phylogenetic signal for the genes examined, both of which can be identified as sources of error in resolving phylogenetic relationships (Hillis, 1998). The most robust strategy for elucidating relationships and gaining further insights into the evolutionary history and patterns of diversification within the Cyprininae is a focused effort to increase taxon sampling across this diverse and important subfamily.

In the past two decades, the mitochondrial 16S rRNA gene has been widely used to explore the phylogenetic relationships of fishes at varying taxonomic levels [e.g. at the order level (Ortí and Meyer, 1997); the familial level (Waters et al., 2000); the subfamily level (Ortí et al., 1996; Harris and Mayden, 2001); the generic level (Moyer et al., 2004); and the species level (Chakraborty and Iwatsuki, 2006; Harris and Mayden, 2001)]. Therefore, the 16S rRNA gene has great potential for the inference of divergences among the cyprinid lineages and resolution of relationships within Cyprininae. Some recent analyses using the mitochondrial 16S rRNA gene, however, have depicted limited resolution of relationships within the Cyprinidae because of low nodal support (Gilles et al., 1998; Simons and Mayden, 1998; Simons et al., 2003; Li et al., 2005). It is well known that phylogenetic inferences depend not only on the appropriate choice of the marker, but also rely on comparative studies of the inherent phylogenetic contents of the molecular markers, such as base compositions, structural characteristics, parameter estimations and the adopted methods (Rosenberg and Kumar, 2003; Telford et al., 2005). One approach to improve the phylogenetic performance of the mitochondrial 16S rRNA gene is to incorporate information regarding the molecular structure of the marker in analyses for more accurate phylogenetic inference.

An important aspect of rRNA genes is that they have conserved secondary structures that are moderately well conserved among distantly related taxa (Caetano-Anollés, 2002). Considering these secondary structural features, rRNA can be divided into paired (stem) and unpaired (loop) regions, and compensatory substitutions occur frequently in the paired regions, a property that contradicts the assumption of independent mutations. Several authors have proposed incorporating differential weighting schemes into parsimony analyses to account for compensatory substitutions in the rRNA genes (Wheeler and Honeycutt, 1988; Springer et al., 1995; Ortí et al., 1996; Bakker et al., 1994; Dixon and Hillis, 1993). However, these weighting schemes have provide nearly equivalent resolu-

tions of relationships relative to equal weighting, largely due to the previous studies not accounting for different selective pressures on stem positions (Wang and Lee, 2002). Recently, simulation studies (Savill et al., 2001) and studies of experimental phylogenies (Higgs, 1998, 2000; Xia et al., 2003; Kjer, 2004; Brown, 2005) had developed likelihood-based methods to evaluate phylogeny and these have resulted in improved phylogenetic resolution. However, these previous studies have mostly focused on lineages diverging across expansive time scales and not on relatively closely related taxa.

Herein, we pursue a detailed analysis of the mitochondrial 16S rRNA gene sequences from the subfamily Cyprininae (sensu Howes, 1991) to shed light on how to best improve the phylogenetic resolution of these fishes, whose origin has dated to no more than 27 MYA estimated from the mitochondrial cytochrome *b* and partial fragments of 16S genes, respectively (Zardoya and Doadrio, 1999; Gilles et al., 1998). The aims of this study are to (1) provide a more inclusive phylogeny within the subfamily Cyprininae; (2) evaluate how to incorporate secondary structural constraints into analyses and improve the information content and performance of the mitochondrial 16S rRNA gene marker in the inference of cyprinine relationships; and (3) to evaluate the resolution of cyprinine phylogeny based on the mitochondrial 16S rRNA gene.

2. Materials and methods

2.1. Samples

For the ingroup mitochondrial 16S rRNA gene sequences were obtained for 84 taxa from the subfamily Cyprininae and nine non-Cyprininae species. Five non-cyprinid species (*Myxocyprinus asiaticus*, *Paramisgurnus dabryanus*, *Micronemacheilus pulcher*, *Misgurnus* sp., and *Pseudogastromyzon fangi*) were selected as outgtoups due to their unambiguous relationships with the family Cyprinidae, relative to the Cyprininae, as determined from both morphological and molecular phylogenetic investigations (Liu and Chen, 2003). New sequences used in this study are listed in Table 1 and have been deposited in GenBank. Species identifications and collections of origin are also given in Table 1. Previously published sequences of *Cyprinus carpio* (NC001606) and *Carassius auratus* (NC002079) were derived from GenBank.

All tissues used for DNA extraction were preserved in 95% ethanol and deposited in the Freshwater Fish Museum of Institute of Hydrobiology of Chinese Academy of Sciences.

2.2. DNA extraction, PCR amplification and sequencing

Total DNA was extracted from muscle or fin tissues using phenol/chloroform extraction procedure (Sambrook et al., 1989). The mtDNA 16S rRNA gene was amplified using standard PCR techniques. Thermal cycle amplifications were performed in 60 μL reaction volumes containing 6 μL 10× buffer, 0.75 mM of each dNTP, 3 μL of each primer, 1.5 U of Taq polymerase (Biostar), and approximately 100 ng of genomic DNA. Amplification proceeded with a primary denaturation step at 95° for 3 min, then 30 cycles of denaturation at 94° for 30 s, annealing at 58°–62° for 30 s and extension at 72° for 60 s, with a final extension of 5 min at 72°. After PCR amplification, a 4 μL sample of each PCR product was detected by 0.8% low-melting agarose gels; the remaining amplified product was purified using the BioStar Glassmilk DNA purification Kit following the manufacturer's protocols and sequenced directly by United Gene Corporation. In this study, two pairs of primers (16Sp1F: 5′-CTT ACA CCG AGA ARA CAT C-3′ and 16Sp1R: 5′-CTT AAG CTC CAA AGG GTC-3′, 16Sp2F: 5′-GAC CTG TAT GAA TGG CTA A-3′ and 16Sp2R: 5′-CTR GGA AGA GGA TTT GAA CC-3′) were used to amplify and sequence the mitochondrial 16S rRNA gene.

2.3. Sequence alignment and partition identification

Previously published sequences were retrieved from the rRNA database (Wuyts et al., 2004; http://www.psb.ugent.be/rRNA) for construction of the complete matrix. The employed in this analysis matrix uses a special distribution format, similar to the sequence files from the rRNA WWW server, to represent knowledge of secondary structure and to be identified by the alignment editor software Seaview (Galtier, 1996). Alignment was refined by eye and saved as a model for subsequent analysis. The orthologous sequences were crudely aligned by Clustal X version 1.8 with default settings (Thompson et al., 1997) and the resulting alignments were refined with reference to the above mentioned model and saved as template for subsequent analyses. To account for the 16S rRNA secondary structure for cyprinids fishes, we used two steps to elucidate structural elements following the methods of Springer and Douzery (1996). First, a criterion that a potential base pairing must occur in at least 75% of the sequences surveyed was used to determine the potential base pairing. Second, we searched for compensatory substitutions as evidence to validate these putative stems.

2.4. Molecular dynamics

Based on secondary structural model described above, base compositions for the different structural categories and their combinations were calculated across all taxa using PAUP*4.0b10 (Swofford, 2003). Chi-square (χ^2) test of base heterogeneity was determined to test for compositional biases existing for the whole gene and gene partitions as implemented in PAUP*4.0b10. To assess phylogenetic signal in the cyprinine MtDNA 16S rRNA gene sequences, the $g1$ skewness statistic (Hillis and Huelsenbeck, 1992) was calculated for 10,000 randomly sampled trees, and permutation tests using 1000 replicates were conducted using PAUP*4.0b10.

Table 1
Samples of species examined in this study, along with voucher specimen catalogue numbers (where available), locality information for specimens, and GenBank accession numbers for the sequenced fragments of 16S rRNA gene

Subfamily	Taxon	Voucher specimen	Sampling location	Accession No.
Cyprininae				
	Cyprinus multitaeniata	IHBCY0308010	Guiping, Guangxi Prov.	DQ845845
	Procypris rabaudi	IHBCY0308496	Hejiang, Sichuan Prov.	DQ845846
	Cyprinus carpio			NC_001606
	Carassius auratus			NC_002079
	Carassius carassius			AY714387
Schizothoracinae				
	Schizothorax myzostomus	IHBCY0605201	Guyong, Yunnan Prov.	DQ845850
	Gymnocypris przewalskii hap1	IHBCY0308689	Qinghai Lake, Qinghai Prov.	DQ845851
	Gymnocypris przewalskii hap2	IHBCY0308795	Qinghai Lake, Qinghai Prov.	DQ845852
	Gymnocypris eckloni hap1	IHBCY0308688	Huanghe, Qinghai Prov.	DQ845853
	Gymnocypris eckloni hap2	IHBCY0308672	Huanghe, Qinghai Prov.	DQ845854
	Platypharodon extremus	IHBCY0308702	Qinghai Lake, Qinghai Prov.	DQ845855
	Schizopygosis pylzovi	IHBCY0308716	Huanghe, Qinghai Prov.	DQ845856
	Schizopygosis younghusbandi	IHBCY0380468	Bomi, Xizang Prov.	DQ845857
	Schizothorax labiatus	IHBCY0380470	Chayu, Xizang Prov.	DQ845858
	Gymnodiptychus dybowskii	IHBCY0405275	Yili, Xinjiang Prov.	DQ845859
	Schizothorax meridionalis	IHBCY0380690	Yingjiang, Yunnan Prov.	DQ845847
	Schizothorax molesworthi	IHBCY0305085	Chayu, Xizang Prov.	DQ845848
	Schizothorax dulongensis	IHBCY0206254	Guyong, Yunnan Prov.	DQ845849
	Ptychobarbus kaznakovi	IHBCY0505344	Chalong, Yunnan Prov.	DQ845916
	Ptychobarbus dipogon	IHBCY0510091	Lasa, Xizang Prov.	DQ845917
	Oxygymnocypris stewartii	IHBCY0510084	Lasa, Xizang Prov.	DQ845918
	Schizothorax prenanti	IHBCY0512001	Ya'an, Sichuan Prov.	DQ845910
	Schizothorax lantsangensis	IHBCY0605028	Baoshan, Yunnan Prov.	DQ845911
	Schizothorax lissolabiatus	IHBCY0605021	Baoshan, Yunnan Prov.	DQ845912
	Gymnocypris potanini	IHBCY0605278	Tengchong, Yunnan Prov.	DQ845897
	Schizothorax argentatus	IHBCY0505299	Tekesi, Xingjiang Prov.	DQ845898
	Schizothorax pseudaksaiensis	IHBCY0505288	Yining, Xingjiang Prov.	DQ845899
	Gymnocypris scolistomus	IHBCY0380715	Shunmucuo , Qinghai Prov.	DQ845903
Barbinae				
	Tor qiaojiensis	IHBCY0205003	Yingjiang, Yunnan Prov.	DQ845873
	barbodes hexagonolepis	IHBCY0410007	Lasa, Xizang Prov.	DQ845874
	Barbodes sp.	IHBCY0130022	Malaysia	DQ845875
	Tor sinensis	IHBCY0308619	Mengla, Yunnan Prov.	DQ845876
	Tor douronensis	IHBCY0405871	Menglun, Yunnan Prov.	DQ845877
	Percocypris pingi	IHBCY0505009	Hejiang, Sichuan Prov.	DQ845878
	Barbaus barbus		France	DQ845879
	Puntius conchorinus	IHBCY0408009	Aquarium, Wuhan	DQ845880
	Barbonymus schwanenfdi	IHBCY0408003	Aquarium, Wuhan	DQ845906
	Sinocyclocheilus yishanensis	IHBCY0410013	Yishan, Guangxi Prov.	DQ845908
	Percocypris retrodorslis	IHBCY0505008	Baoshan, Yunnan Prov.	DQ845909
	Acrossocheilus beijiangensis	IHBCY0403416	Rong'an, Guangxi Prov.	DQ845869
	Barbodes vernayi	IHBCY0405871	Mengla, Yunnan Prov.	DQ845870
	Mystacoleucus lepturus	IHBCY0405396	Mengla, Yunnan Prov.	DQ845871
	Mystacoleucus marginatus	IHBCY0411002	Mengla, Yunnan Prov.	DQ845913
	Sinocyclocheilus tingi	IHBCY0210313	Fuxian Lake, Yunnan Prov.	DQ845866
	Acrossocheilus sp.	IHBCY0405415	Shanghang, Fujian Prov.	DQ845868
	Acrossocheilus hemspinus	IHBCY0403452	Rong'an, Guangxi Prov.	DQ845867
	Sinocyclocheilus jii	IHBCY0380717	Yishan, Guangxi Prov.	DQ845923
	Sinocyclocheilus grahami	IHBCY0410014	Dianchi, Yunnan Prov.	DQ845924
	Sinocyclocheilus macrolepiss	IHBCY0410006	Libo, Guizhou Prov.	DQ845925
	Sinocyclocheilus macroscalus	IHBCY0410011	Liuliang, Yunnan Prov.	DQ845927
	Sinocyclocheilus yangzongensis	IHBCY0410008	Yangzonghai, Yunnan Prov.	DQ845926
	Barbus sp.	AFR292	Africa	DQ845860
	Onychostoma sima	IHBCY0306001	Hejiang, Sichuan Prov.	DQ845861
	Onychostoma gerlachi	IHBCY0405405	Jinghong, Yunnan Prov.	DQ845862
	Hampala macrolepidota	IHBCY0405393	Mengla, Yunnan Prov.	DQ845863
	Spinibarbus sinensis	IHBCY0403005	Huangshan, Anhui Prov.	DQ845864
	Spinibarbus hollandi	IHBCY0205001	Tunxi, Anhui Prov.	DQ845865
	Sikukia stejnegeri	IHBCY0405381	Menglun, Yunnan Prov.	DQ845872

(*continued on next page*)

Table 1 (*continued*)

Subfamily	Taxon	Voucher specimen	Sampling location	Accession No.
Labeoninae				
	Sinilabeo laticeps	IHBCY0210030	Mengla, Yunnan Prov.	DQ845904
	Epalzeorhynchus frenatus	IHBCY0408007	Aquarium, Yunnan Prov.	DQ845905
	Garra mirofrontis	IHBCY0405433	Menglun, Yunnan Prov.	DQ845907
	Garra orientalis	IHBCY0403441	Ledong, Hainan Prov.	DQ845884
	Garra kempi	IHBCY0309091	Chayu, Xizang Prov.	DQ845885
	Semilabeo notabilis	IHBCY0405392	Jinxiu, Guangxi Prov.	DQ845886
	Parasinilabeo assimilis	IHBCY0308002	Rong'an, Guangxi Prov.	DQ845887
	Discogobio tetrabarbatus	IHBCY0308003	Rong'an, Guangxi Prov.	DQ845888
	Discogobio laticeps	IHBCY0308942	Tain'e, Guangxi Prov.	DQ845889
	Labiobarbus lineatus	IHBCY0407001	Mengla, Yunnan Prov.	DQ845914
	Placocheilus crytonemus	IHBCY0504726	Tian'e, Guangxi Prov.	DQ845915
	Discogobio bismargaritus hap2	IHBCY0308734	Liuzhou, Guangxi Prov.	DQ845890
	Rectoris posehensis	IHBCY0204016	Luxi, Hunan Prov.	DQ845891
	Osteochilus salsburyi	IHBCY0308001	Rong'an, Guangxi Prov.	DQ845892
	Epalzeorhynchus bicornis	IHBCY0505291	Liuku, Yunnan Prov.	DQ845919
	Ptychidio jordani	IHBCY0308004	Tian'e, Guangxi Prov.	DQ845893
	Pseudogyrinocheilus procheilus	IHBCY0405017	Panzhihua, Sichuan Prov.	DQ845894
	Pseudocrossocheilus bamaensis	IHBCY0509003	Tian'e, Guangxi Prov.	DQ845895
	Labeo yunnanensis	IHBCY0301133	Mengla, Yunnan Prov.	DQ845881
	Crossocheilus latius	IHBCY0308005	Menglun, Yunnan Prov.	DQ845882
	Cirrhinus molitorella	IHBCY0308009	Tengxian, Guangxi Prov.	DQ845883
	Discogobio bismargaritus hap1	IHBCY0308821	Boyi, Yunnan Prov.	DQ845900
	Discogobio sp.	IHBCY0308882	Jinxiu, Guangxi Prov.	DQ845901
	Lobocheilus melanotaenia	IHBCY0405266	Menglun, Yunnan Prov.	DQ845902
Other subfamily				
	Gobio gobio		France	DQ845928
	Gobiobotia abbreviate	IHBCY0303091	Tian'e, Guangxi Prov.	DQ845929
	Squaliobarbus curriculus	IHBCY0407001	Jinkou, Hubei Prov.	DQ845930
	Xenocypris argentea	NRMT2281	Taoyuan, Hunan Prov.	DQ845931
	Hypophthalmichthys molitrier	IHBCY0380497	Chunxi, Hunan Prov.	DQ845932
	Ctenopharyngodon idellus	IHBCY0380488	Hengxian, Guangxi Prov.	DQ845933
	Hemiculter bleekeri	IHBCY0380491	Wuhan, Hubei Prov.	DQ845934
	Leuciscus rutilus			AF038484
	Notropis texanus			AY216552
	Abramis bjoerkna			AF038484
	Danio rerio			NC_002333
Outgroups				
	Misgurnus sp.	IHBCY0506227	Panzhihua, Sichuan Prov.	DQ845935
	Myxocyprinus asiaticus	IHBCY0305001	Wuhan, Hubei Prov.	DQ845896
	Micronemacheilus pulcher	IHBCY0380485	Rong'an, Guangxi Prov.	DQ845921
	Paramisgurnus dabryanus	IHBCY0380486	Rong'an, Guangxi Prov.	DQ845922
	Pseudogastromyzon fangi	IHBCY0380487	Hengxian, Guangxi Prov.	DQ845920

Species identifications follow those proposed by Chen (1998).

Under the assumption of maximum parsimony (Swofford et al., 1996), both substitution rates and the explicit number of changes per site were estimated by tracing changes on the shortest resolved tree (referred to the phylogenetic results of the current study) using MacClade 4 (Maddison and Maddison, 1992). Following the method described by Vawter and Brown (1993), the relative rate of each kind of nucleotide substitution for each structural category, as well as for the entire gene, was corrected for base composition. The site-to-site rate variation was performed by PAML3.14 following the protocol of Yang and Kumar (1996).

For stem regions, to investigate the possible effect of secondary structural constraints on phylogenetic inference, a tally of the transformations maintaining or disrupting the pairing in double-stranded regions, including single changes and double changes were compared with expected values (Dixon and Hillis, 1993). To account for the degree of independence of substitutions occurring in the stem characters a relative weighting, suggested by Dixon and Hillis (1993), was estimated.

2.5. Phylogenetic analyses

Maximum parsimony (MP) and likelihood-based methods were carried out to investigate evolutionary relationships within the Cyprininae. MP analysis was conducted in PAUP*4.0b10 (Swofford, 2003), while the likelihood-based methods were completed using MrBayes3.0 (Ronquist and Huelsenbeck, 2003) and PHASE 1.1 (Jow et al., 2002), separately. The MP analyses were conducted follow-

ing different weighting schemes: (1) equal weights for all changes; (2) several forms of loop–stem weighting, such as 1:0.8 (Dixon and Hillis, 1993), 1:0.66 (this study), 1:0.6 (Springer and Douzery, 1996) and 1:0.5 (strict dependence of nucleotide in stem regions), 1:0 (loops only) and 0:1 (stems only). All MP trees were constructed using a heuristic search with 50 random, stepwise additions of taxa and tree bisection–reconnection (TBR) branch swapping. Quantitative support for recovered nodes was estimated using a non-parametric bootstrap analysis with 1000 pseudoreplicates and a heuristic search with 10 addition sequences replicates.

Bayesian analysis permits efficient searching of parameter space for complex likelihood models, all of which can be applied to different partitions of a data set (Huelsenbeck and Ronquist, 2001). Thus, we use different substitution for the paired and unpaired regions. For unpaired partitions the GTR + I + G model, as determined by Modeltest version 3.06 (Posada and Crandall, 1998), was employed. For the paired regions, those probabilistic models considered the process of substitution on 16 characters formed by 2 nucleotide residues paired were employed. In MrBayes3.0, there is only a double model responsible for nucleotide substitutions of paired region, while there are more models in PHASE. Because of the complexity of testing each of PHASE models, the pertinent models including RNA7A and RNA16A were prior to be used. To find out how base covariation of stem regions affects the phylogenetic performance of the gene sequences, we also conducted Bayesian analysis with the uniform GTR + I + G model. Using the Markov Chain Monte Carlo (MCMC) algorithm, the independent runs were performed with default settings, each with four separate chains (three hot, one cold) and 3,000,000 generations, sampling every 500 generations. Graphical inspection of the $-\ln L$ included in "sump" revealed that stationary was reached with the "burn-in" After discarding the burn-in, posterior probabilities (PP) for topologies were then assessed. To test consistency of the results, the analyses were repeated three times using different starting trees. The Bayesian framework employed in PHASE used random starting trees and the parameters of the substitution models were estimated during the analysis. Following Hudelot et al. (2003), we adopted a conservative burn-in period and used 2,000,000 initial generations. Later, another 1,000,000 generations were run with sampling every 500 generations. A consensus tree with branch lengths and nodal PP support was generated using PHYLIP (Felsenstein, 1989).

2.6. Tests of alternative topologies

To test the effects of different loop–stem weighting schemes on the parsimony-based estimation of relationships of Cyprininae, different MP topologies recovered in the present study were compared using the Wilcoxon signed-ranks, two-tailed probability test (Templeton, 1983; Felsenstein, 1985). In addition, competing topologies were compared using the SH-test (Shimodaira and Hasegawa, 1999) implemented in PAUP*4.0b10 to test for significant differences in tree length. This test was performed using RELL with 1000 bootstrap replicates; the results were evaluated as a one-tailed test.

2.7. Variable length bootstrap

We employed variable length bootstrap analysis in order to investigate the relationship between improvement of the phylogenetic resolution and lengths of mitochondrial 16S rRNA gene sequences. In this analysis, bootstrap support is estimated as a function of a variable number of resampled characters (Springer et al., 1999), and nucleotide sites were resampled to generate bootstrap pseudomatrices of extension from 2000 to 30,000 characters with increasing steps of 2000 sites. All bootstrap searches were then performed using MP analyses with PAUP*4.0b10.

3. Results

3.1. Sequence characteristics

Relative to the secondary structure of common carp (*C. carpio*) available in the European ribosomal RNA database (http://www.psb.ugent.be/rRNA), approximately 10% of the mitochondrial 16S rRNA helices in our model have been refined. The alignment to secondary structure across all taxa was 1791 bases, of which one region (sites 449–494) was deleted from our analyses because of unreliable alignment. Nucleotide frequencies in cyprinid 16S rRNA gene sequences are illustrated in Fig. 1. In our comparisons of taxa the mitochondrial fragments of Cyprininae have an excess of Adenine (A) relative to other nucleotides, a pattern observed in other fishes (Moyer et al., 2004). The 16S rRNA gene displays similar base compositions of thymine (T), cytosine (C), and guanine (G) across the cyprinids, similar to the results outlined by Liu (2004) for the mitochondrial 12S rRNA gene for cypriniform fishes. Nucleotide compositions of both stems and loops of the

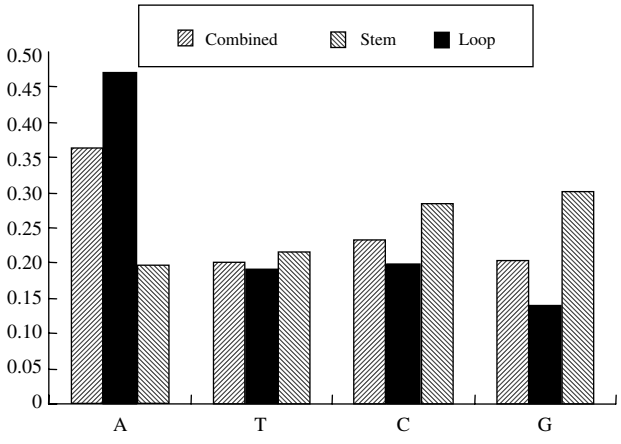

Fig. 1. Mean base composition of 16S rRNA gene sequences among the taxa analyzed.

16S rRNA gene are provided in Fig. 1. For the 16S rRNA gene loop regions display a higher A + T bias (66.14%) than the whole gene due to an extreme A-bias (47.07%) and G is the least represented base. However, stem regions display a higher G + C bias (58.72%), of which the proportion of G (30.24%) is slightly higher than that of C. Chi-square tests of homogeneous base frequencies among all taxa in stem regions ($\chi^2 = 23.923595$, df = 288, $P = 1.00$), in loop regions ($\chi^2 = 214.194783$, df = 288, $P = 0.99$), and overall ($\chi^2 = 157.299373$, df = 288, $P = 1.00$) failed to reject the null hypothesis, indicating that base compositions are stationary across all taxa surveyed.

3.2. Patterns of nucleotide change

The relative nucleotide substitutions rates in stems, loops, and the overall sequences are shown in Fig. 2a. The overall substitution rate in loops is two times higher than that in stems. Taking transformations into account separately, the transition rate in loops is approximate 80% higher than that in stems, whereas the transversion rate is fivefold greater. It is well known that a transition–transversion rate bias has been observed in most mitochondrial genes (Sloss et al., 2004). As expected, there is an obvious transition–transversion rate bias, being 5.78, 1.78, and 2.59 for stem, loop, and their combinations, respectively. However, the proportions of A ↔ G and C ↔ T differ across the structural categories and the overall sequences (Fig. 2b). The C ↔ T transitions are the main type of change in loop regions and the entire data set, whereas A ↔ G transitions are more abundant in stem regions. Although the assumption of equal transversions probabilities across the structural categories and entire data set was also refuted, uniform proportions were found between loops and overall sequences.

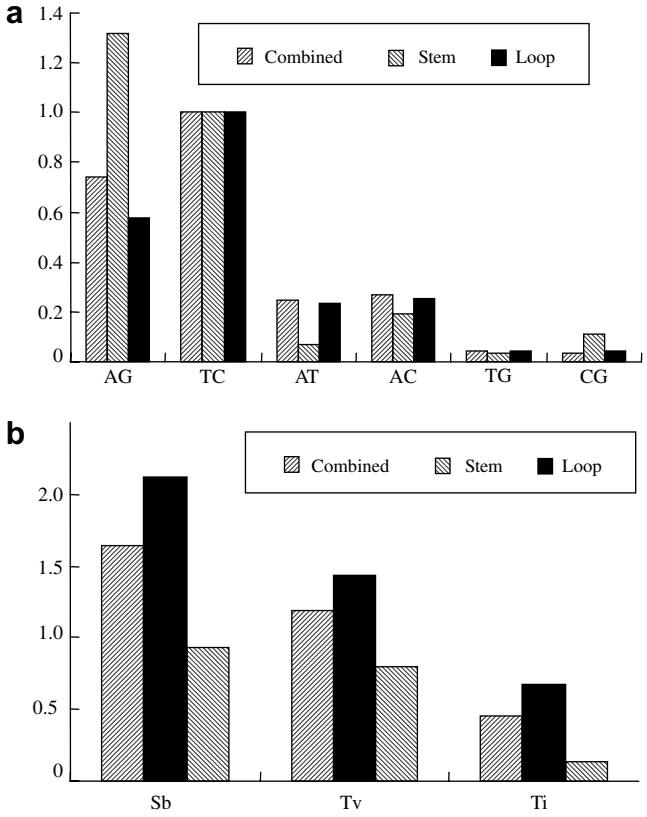

Fig. 2. Comparison of maximum parsimony estimates referred to the topology from MCMC algorithm under the RNA16A model. (a) Relative rates of the different types of base changes (corrected for base composition) in 16S rRNA gene sequences. For the whole gene and each structural class, rates are expressed as proportion of the number of C ↔ T changes; (b) rates of nucleotide substitutions, transversions and transitions in the whole fragments and each structural class of 16S rRNA gene sequences (Sb = substitutions per site, Tv = transversions per site, and Ti = transitions per site).

Lacking rate heterogeneity among sites, the distributions of the number of substitutions per site must follow a Poisson distribution. This assumption is tested for different structural regions and for the overall sequences of the cyprinid 16S rRNA gene sequences using comparisons between the observed and the expected numbers of substitutions. Results showed that the number of invariant and highly variable sites significantly greater than expected, but those with intermediate numbers of substitutions underestimated (Table 2). Assuming a Gamma distribution of rates across sites, the shape parameter α was estimated to be 0.5247, using PAML with reference to the reconstructed phylogeny.

3.3. Compensatory base changes and positional weighting

The observed and expected numbers of compensatory mutations in stem regions of the MtDNA 16S rRNA gene are shown in Table 3. There are significantly more compensatory mutations than expected ($\chi^2 = 900.3$ and $\chi^2 = 952.2$, df = 1, $P < 0.0001$, for single and double substitutions, respectively). Of all substitutions occurring in stem regions, 1116 of 1548 were observed to maintain base pairings, but only 301.1 compensatory mutations were expected. These results indicate that at least in cyprinid fishes selection for compensatory substitutions is strong and an appropriate relative weighting to account for this mode of evolution in stem regions and their non-independence is critical for accurate an phylogenetic analysis. According to the suggestion of Dixon and Hillis (1993), an exact weighting for stem regions should be 0.66. This value is less than what Dixon and Hillis (1993) suggested for 28S rRNA and equal to that for 12S rRNA (Wang and Lee, 2002) and for partial 18S rRNA gene sequences (Bakker et al., 1994).

3.4. Phylogenetic analysis

3.4.1. Maximum parsimony

The unweighted MP analysis resulted in 958 equally parsimonious trees; a strict consensus of these trees is provided in Fig. 3, where bootstrap values are indicated at nodes with >50% support. Monophyly of the family Cyprinidae was recovered with strong bootstrap support (95%), whereas Monophyly of the nominate subfamily Cyprininae

Table 2
Observed and expected substitution values for stem, loop, and the across entire 16S rRNA gene in Cyprininae

Number of changes	Stem		Loop		Overall	
	Observed	Expected	Observed	Expected	Observed	Expected
0	420	213.85	532	468.6	953	407.9
1	69	224.5	92	302.66	159	530.87
2	39	117.81	69	61.21	105	345.47
3	18	41.23	43	3.4	62	149.88
4	23	10.82	30	0.44	50	48.77
5	15	2.7	30	0.05	41	12.69
6	5	0.4	22	0	32	2.75
7	6	0.06	29	0	37	0.51
8	6	0.01	16	0	19	0.08
9	6	0	18	0	22	0.01
10	4	0	0	0	0	0

Observed values were derived from mapping substitutions on the phylogeny estimated with the likelihood-based method under the RNA16A model. Expected numbers of substitutions are derived from a Poisson distribution. Chi-square goodness-of-fit tests show significance ($P = 0.001$) for stems ($\chi^2 = 447.9$), loops ($\chi^2 = 1432.8$), and the whole sites ($\chi^2 = 3247.8$).

Table 3
Types of substitutions observed in stem regions of the 16S rRNA gene sequences of Cyprininae

Type of substitution	Observed	Expected	Chi-square tests
Single			
Base pairing to base pairing	90.5	358	
Base pairing to non-base pairing	633.5	366	$\chi^2 = 900.3^*$
Double			
Base pairing to base pairing	105.3	379	
Base pairing to non-base pairing	306.7	33	$\chi^2 = 952.2^*$

Significantly ($P < 0.05$) different topologies are indicated by asterisks.

was supported with somewhat low bootstrap value (59%). Within the subfamily Cyprininae, relationships among major lineages were not adequately resolved.

The trees resulting from parsimony analyses using different down-weightings for structural constraints are mostly congruent in topology with the tree (Fig. 3) except for a few topological differences that were not strongly supported. The relative down-weighting in stems reduced the number of MPTs and increased phylogenetic resolution. However, despite the topology imposed with the 1:0.66 weighting for loop–stem regions was determined to be the best one by both unweighted MP Templeton and winning-sites tests, the relationships among the major lineages of the Cyprininae remained unsolved.

When only loop or stem characters were used in MP analyses, the monophyly of the Cyprininae could not be recovered and the topologies resolved showed significant differences with unweighted MP and the other weighted MP trees.

3.4.2. Bayesian analyses

SH test revealed that Bayesian analysis under the 16-state model performed best (Table 4). However, it is important to note that topologies under all surveyed models were virtually indistinguishable except for posterior probabilities (PP) on some branches from the above replicates which deviated more than 10% (Fig. 4). In Bayesian tree, both the family Cyprinidae and the subfamily Cyprininae were found to be monophyletic with strong posterior probability support (PP = 100). Within the Cyprininae, *Procypris rabaudi* was recovered as the sister taxon to all other Cyprinines with a high posterior probability (PP = 100), followed by the genera *Sinocyclocheilus* + *Cyprinus* as a monophyletic group (PP = 86). Strong support was also found for a group (PP = 100) including tribe Labeonini, a paraphyletic genus *Tor* with respect to *Barbodes hexagonolepis* and *Barbodes* sp. An additional Schizothoracin clade was strongly supported (PP = 100); within this clade the sister group relationship between the genera *Schizothorax* and *Percocypris* was strongly supported (PP = 100).

3.5. Effect of molecular sequence length

According to the variable length bootstrap curves, the number of nodes achieving at least 50% bootstrap support increases with the resampling of more sites (Fig. 5). Improvement is significant for the number of nodes attaining at least 90% support, which mainly occurred in more recent nodes. However, little further increase in bootstrap values was found with the resampling of sites exceeding 30,000 bp (not shown). When support values for individual nodes were taken into account (Fig. 6; letters A–H refer to nodes recovered in MP), nodes A and B achieved 90% bootstrap support with 9000 bp, and node G with 24,000 bp. Nodes D, E, and H achieved moderate support with 30,000 bp. Nodes F and I achieved less than 50% bootstrap support, even when resampling of sites was 100,000 bp (not shown).

4. Discussions

4.1. Molecular dynamics and comparative method implications

A better understanding of the molecular dynamics of DNA sequences, such as base compositions, nucleotide

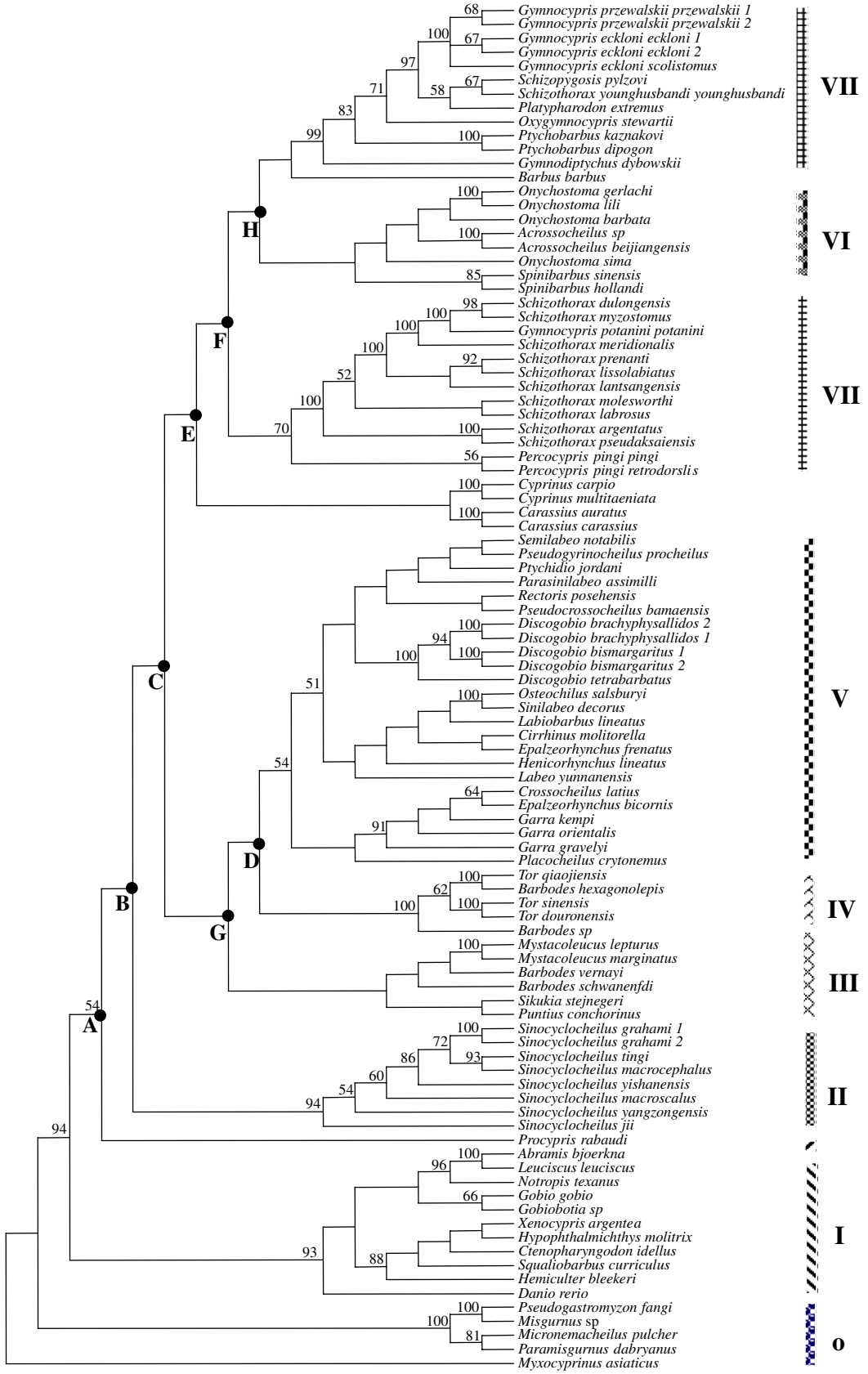

Fig. 3. Strict consensus of 958 trees results from unweighted maximum parsimony analysis of the mitochondrial 16S rRNA gene sequences (length = 4313, CI = 0.3144, RI = 0.6277). Numbers at nodes represent non-parametric bootstrap values >50%. O = outgroup, I = Leuciscinae, II = *Sinocyclocheilus* + *Cyprinus*, III = *Carassius* + *Puntius conchorinus* + *Sikukia stejnegeri* + *Barbodes schwanenfdi* + *Barbodes vernayi* + *Mystacoleucus*, IV = *Tor* + *Barbodes* sp., V = Labeonini, VII = Schizothoractin, and VIII = *Onychostoma* + *Acrossocheilus*).

Table 4
Templeton and winning-sites tests of the weighted trees of Cyprininae generated with a series of stem–loop weightings

Weighted	Tree length	Templeton		Winning-sites		HI	MPTs
		Ts	*P*	Counts	*P*		
1:1	4313	4866.5	0.6455	72 (−70)	0.9331	0.6856	958
0.8:1	4310	4268.0	0.8910	65 (−66)	1.0000	0.6887	163
0.66:1	4304	(Best)				0.6905	18
0.6:1	4321	7722.5	0.2974	101 (−82)	0.1833	0.6916	36
0.5:1	4328	3159.0	0.0779	66 (−57)	0.4707	0.6933	18
0:1	4769	9397.5	<0.0001*	268 (−68)	<0.0001*	0.7020	111
1:0	4403	6873.0	<0.0001*	129 (−74)	0.0002*	0.5732	1500

HI, homoplasy index; MPTs, number of most parsimonious trees. Significantly ($P > 0.05$) different topologies are indicated by asterisks.

substitution patterns, and rate variations among sites, should improve the phylogenetic performance of the DNA sequences (Yang, 1994; Yang and Kumar, 1996). Of these variables, two key factors include the functional constraint of maintaining base pairing on stem regions and rate variation across sites for the mitochondrial 16S rRNA gene sequences.

Because of deviations from the assumption of character independence in phylogenetic analyses, positional weighting has been recommended as a method to correct for compensatory substitutions occurring in paired regions (Dixon and Hillis, 1993; Ortí et al., 1996; Wang and Lee, 2002). In this study, the down-weighting on non-independent substitution of the MtDNA 16S r RNA gene sequences has little impact on the tree topology, even when the loop–stem weighting of 1:0.66 was used. We envision three reasons for the limited impact of the weighting schema on the stem regions for phylogenetic inference in Cyprininae. First, the accumulated phylogenetic information in the stem regions lacked enough information content to resolve relationships at lower taxonomic levels. The poorly established cyprinine relationships based on the phylogenetic information in stem regions 16S rRNA gene validates this tendency (not shown). Second, the effect of selective pressures on parsimony analysis is loose (Van de Peer, 1993; Sloss et al., 2004). Finally, the down-weighting schemes of non-independent characters are too simplified and do not elucidate the different selective pressures in stem regions (Telford et al., 2005), and thus are not adequate for real data sets. Therefore, our results suggest that the down-weighting of characters of stems relative to loops might not be helpful to improve phylogenetic inference.

Compared with the basic assumptions of parsimony analysis, the model-based Bayesian analyses generally should be preferred, in that it can accommodate better estimates of appropriate models to deal with the functional constraints of maintaining base pairing on stem regions (Hudelot et al., 2003; Kjer, 2004; Brown, 2005). Likelihood-based phylogenetic analyses can be demonstrated to be more effective in improving the phylogenetic performance than weighted parsimony analysis (Table 5). In Bayesian analyses, the phylogenetic performance under the prior super models including of 7-state model, 16-state model, and double model implemented in MrBayes3.0 were all better than observed under a simple GTR + G + I model, an result that provides overwhelming evidence for considering the secondary structural constraints inherent in the rRNA gene to increase phylogenetic resolution. Like the phylogenetic utility of rRNA gene sequences for resolving relationships at deep taxonomic levels, the superiority of model-based approaches in accounting for constraints inherent in secondary structure in more recently diverged taxa can also result in improved resolutions of phylogenetic trees. However, with and without considering the 16S rRNA secondary structure received the almost same topology, and this result indicated that the degree of violation of the assumption of independence of sites is not important as expected and can hardly affect topology of recent divergence phylogeny. At higher taxonomic levels (subfamily and tribe), Bayesian analyses under the double models resulted in more robust support for nodes than did analyses that incorporate the single GTR model. However, the more complex Bayesian models also resulted in unacceptably low support for relationships among closely related taxa (species and genus). An explanation for this scheme is that paired regions may be regarded as being more suitable for distantly related organisms rather than recent divergence.

Rate variation among sites is widespread in the mitochondrial 16S rRNA gene sequences (Whitfield and Cameron, 1998; Misof et al., 2002; Brown, 2005) and should receive more attention in the development of algorithms or models to better explain this pattern of divergence within a gene (Susko et al., 2003). Theoretically, it has been assumed that the distribution of change by site followed a Poisson model (Rzhetsky, 1995). In this study, however, the degree of site-to-site rate variation existed across sequences and displayed a significant deviation from the expected Poisson distribution ($P < 0.001$). From a likelihood-based perspective, the Gamma distribution, used to measure the degree of rate variation among sites (Yang, 1996), also supported the hypothesis that most sites have very high substitution rates. Simultaneously, the proportion of invariable sites is 53.15%, indicating that across the gene more sites are virtually invariable. Thus, the likelihood-based models with rate variation among sites and proportion of invariant sites serve as a potential tool to effectively account for situations of this nature (Yang, 1996).

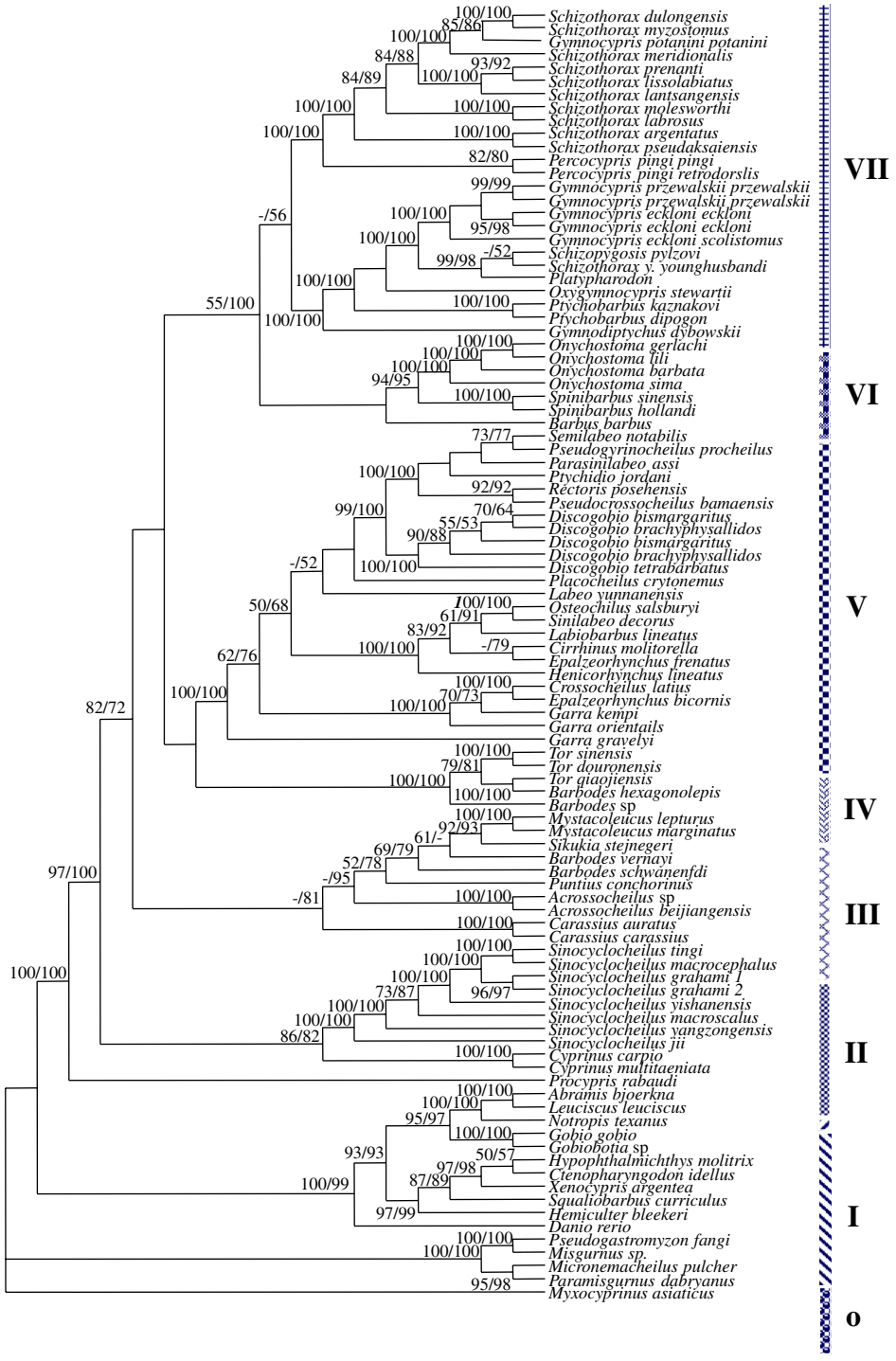

Fig. 4. Phylogenetic relationships of the Cyprininae as determined by Bayesian analysis of 16S rRNA gene sequences incorporating different models. Numbers above or below branches are posterior probabilities values (unconstrained Bayesian: 3201; constrained Bayesian: 3026). O = outgroup, I = Leuciscinae, II = *Sinocyclocheilus* + *Cyprinus*, III = *Carassius* + *Puntius conchorinus* + *Sikukia stejnegeri* + *Barbodes schwanenfdi* + *Barbodes vernayi* + *Mystacoleucus*, IV = *Tor* + *Barbodes* sp., V = Labeonini, VII = Schizothoractin, and VIII = *Onychostoma* + *Acrossocheilus* + *Spinibarbus*).

4.2. Phylogeny

This study presents the first view on the phylogeny of the subfamily Cyprininae using molecular data with comprehensive sampling of the diversity within this lineage. In accordance with hypotheses based on conventional morphological characters, all analyses were congruent and supported the monophyly of the subfamily Cyprininae.

Our analyses are also in agreement with some of the previous morphological studies concerning the monophyly of several main clades within the subfamily Cyprininae. Examining only African Labeins, Reid (1982) defined the

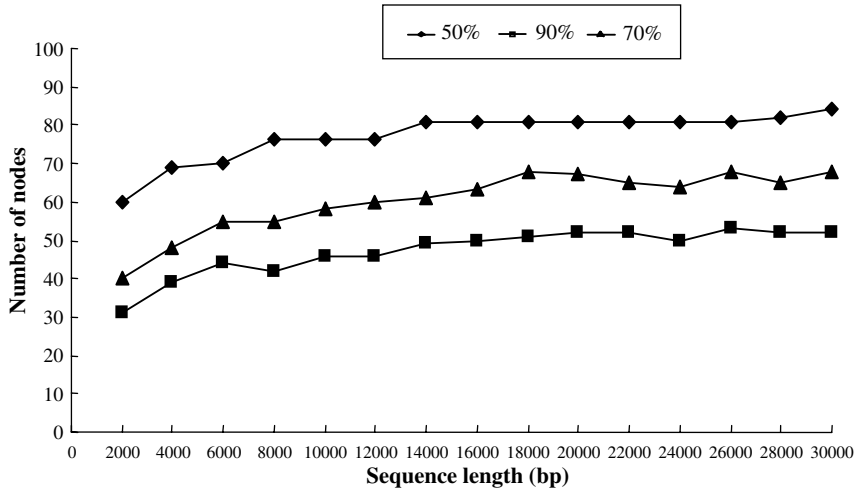

Fig. 5. The plot of the sequence length and the number of nodes achieved the bootstrap value of at least 50%, 70%, and 90% using the variable length bootstrap.

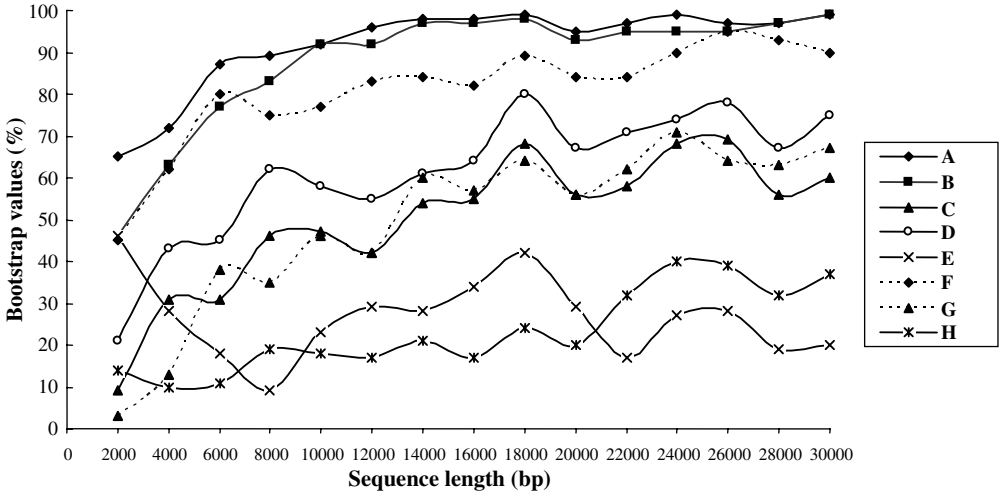

Fig. 6. Plot of 16S rRNA gene sequence length and bootstrap values attained for nodes A–H in the inferred maximum parsimony tree using variable length bootstrap.

Table 5
Shimodaira–Hasegawa test of all topologies of Cyprininae generated by different methods of analysis

Trees	$-\ln L$	Diff. in $-\ln L$	P
Unweighted MP	−21993.76111	45.41847	0.139
Unconstrained Bayesian (GTR + G + I)	−21965.27586	16.93321	0.629
Constrained Bayesian (double model)	−21959.73766	11.39502	0.713
Constrained Bayesian (RNA7A)	−21953.88259	5.53994	0.852
Constrained Bayesian (RNA16A)	−21948.34265	Best	Best

Labein fishes in terms of four derived features: configuration of the parasphenoid, basioccipital process, and a simple last dorsal-fin ray. Chen et al. (1984) also diagnosed tribe Labeonini as a monophyletic group. As expected, we found unequivocal support for the monophyly of the tribe Labeonini, despite only moderate support values from non-parametric bootstrapping.

Our analysis disagrees with traditional cyprinine systematics in many aspects. The monophyly of the currently recognized tribes, such as Schizothoractin, Barbin, and *Cyprinion–Onychostoma* lineages, was rejected in the present analysis. Based on morphological evaluations, Cao et al. (1981) recognized three general grade groups within Schizothoractins. These have been traditionally referred

to as primitive, specialized, and highly specialized schizothoracine fishes, respectively. In this study, Schizothoractins were clearly divided into two subgroups, one containing the genera *Percocypris* and *Schizothorax* and the other represented by the specialized and highly specialized schizothoractin groups of Cao et al. (1981). Based on karyotype data, Yu et al. (1987) proposed that *Percocypris* was closely related to *Schizothorax,* regardless of external morphological similarities or differences. This relationship was strongly supported in the present analysis.

Another area in which our analyses conflict with traditional views is with the monophyly of the genus *Onychostoma* and its relationship to other taxa. The genus is paraphyletic and with the genus *Acrossocheilus*. Based on detailed anatomical studies, Chen (1998) suggested that *Onychostoma* and *Scaphesthes* formed a monophyletic group with a close relationship to *Cyprinion*. However, the lacking the evaluation of character polarity of morphological characters, the suggested dendrogram could not account for the relationships between *Onychostoma* and *Acrossocheilus*. Zhang et al. (pers. comm.) reported that the classification *Acrossocheilus* was chaotic because of the lack of character independence in the morphological data. Therefore, it is imperative future studies include extensive taxonomic sampling of *Acrossocheilus* and potentially related taxa and an integration of molecular data with traditional morphological characters to resolve this systematic problem.

Finally, phylogenetic analysis also revealed that the genus *Garra* is paraphyletic with respect to *Epalzeorhynchus bicornis* and *Crossocheilus latius*, *Tor* is paraphyletic with respect to *barbodes hexagonolepis* and *Barbodes* is paraphyly or polyphyly.

4.3. Phylogenetic signal

Our results suggested that the mitochondrial 16S rRNA data provide limited ability to recover relatively old and long terminal branches in cyprinine phylogeny. Despite efforts to improve the phylogenetic resolution with optimization of likelihood models and parsimony weightings for the structural and functional regions (stem and loop regions), the relationships among genera or tribes within Cyprininae remain unresolved or are not robustly supported. For example, only weak support was found for a sister-group relationship between Labeonini and a clade including a schizothoracin lineage, *Percocypris*, *Spinibarbus*, *Onychostoma*, and *Barbus barbus*, even when the analyzed sequences were generated to lengths longer than 30,000 bp using the variable length bootstrap. This same grouping of taxa had robust support in analyses by Wang et al. (2007) based on the recombination activating gene 2 (RAG2). One possible interpretation for this observation is a low phylogenetic signal for the 16S rRNA gene and a lack of adequate phylogenetic information accumulated to recover areas of the relatively deep phylogeny. On the other hand, relationships among genera within the Labeonini remain unresolved likely due to a combination of short internodal and long terminal branches ("long-branch attraction", Hendy and Penny, 1989) because of a period of rapid radiation events. The estimated divergence times indicated the divergence between Labeonini and Cyprinin occurred 15–19 MYA and the Labeonini experienced a radiation starting about 10 MYA (Wang et al., 2007). Therefore, we can conclude that the mitochondrial 16S rRNA sequences might not contain enough phylogenetic signals for cyprinine divergences that occurred 10–19 MYA.

Interestingly, contradicting evidence for this gene and its phylogenetic utility comes from its applications among Hymenoptera (Simon et al., 1994; Whitfield and Cameron, 1998). In these studies the mitochondrial 16S rRNA provided better resolution of intrageneric relationships, such as in *Schizothorax* and *Sinocyclocheilus*. He and Chen (2006) estimated the adaptive radiation for *Schizothorax* as occurring about 8 MYA based on sequence variation of cytochrome *b*, and an investigation using both cytochrome *b* and ND4 indicated that the *Sinocyclocheilus* lineage experienced a radiation in Mid-Pliocene (about 3.1–4 MYA) (Xiao et al., 2005). These results in these taxonomic groups support the hypothesis that the mitochondrial 16S rRNA gene is an effective tool to trace evolutionary histories spanning the recent 8 MY. Some previous studies have suggested that phylogenies inferred from single genes usually attained insufficient resolution (Nickrent et al., 2000; Philippe, 2000). A reasonable alternative method is combining multiple genes or gene regions into a framework to provide improved resolution and accuracy of our phylogenetic inference (Galewski et al., 2006; Bossuyt et al., 2006) as has recently been done by Mayden et al. (2007) for Cypriniformes. Thus, while in some instances it may be difficult, further phylogenetic studies should make every effort to use extensive sampling of taxa to "break up" long branches that can lead to the phenomenon of long-branch attraction and sequences from multiple mtDNA and nDNA genes should be used to accumulate enough phylogenetic information to resolve the diverged lineage. As outlined by Hillis (1998) and Mayden et al. (2007) increased sampling of either taxa or genes can result in increased accuracy of phylogenetic inference; however, a combined effort is most effective. To accomplish such an effort we encourage researchers to collaborate more worlds wide in the acquisition of taxa and efficiency of gene sequences; only through such a larger-scale collaborative effort can one develop enough samples and data to enhance the confidence in the phylogenetic resolutions.

Acknowledgments

We are grateful to the following collaborators for providing specimens or tissues for this and future studies of cyprinine relationships: Dr. H. Liu, K. Zhao, D. He, Z. Peng, and Z. Chen. We thank Dr. J. Luo, Z. Peng, and W. Wang for providing suggestions and modifications.

We also extend our gratitude to the anonymous reviewers for their useful suggestions. This study was supported by the grants from National Natural Science Foundation of China (NSFC) 30225008, 30300036, and 30530120, and the National Science Foundation Assembling the Tree of Life Program of the USA (NSF EF-0431326).

References

Bakker, T., Olsen, L., Stam, T., Hoek, V.D., 1994. The cladophora complex (chlorophyta): new views based on 18S rRNA gene sequences. Mol. Phylogenet. Evol. 3, 365–382.

Banarescu, P., Coad, B.W., 1991. Cyprinids of Eurasia. In: Winfield, I.J., Nelson, J.S. (Eds.), Cyprinid Fishes. Systematics, Biology and Exploitation. Chapman & Hall, London, pp. 127–155.

Bossuyt, F., Brown, R.F., Hillis, D.M., Cannatella, D.C., Milinkovitch, M.C., 2006. Phylogeny and biogeography of a cosmopolitan frog radiation: late cretaceous diversification resulted in continent-scale endemism in the family Ranidae. Syst. Biol. 55, 579–594.

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

Brown, R.P., 2005. Large subunit mitochondrial rRNA secondary structures and site-specific rate variation in two lizard lineages. J. Mol. Evol. 60, 45–56.

Caetano-Anollés, G., 2002. Tracing the evolution of RNA structure in ribosomes. Nucleic Acids Res. 30, 2575–2587.

Cao, W., Chen, Y., Wu, Y., 1981. In: The Team of the Comprehensive Scientific Expedition to the Qinghai-Xizang Plateau, Chinese Academy of Sciences (Ed.). Origin and Evolution of Schizothoracine Fishes in Relation to the Upheaval of the Xizang Plateau, in Collection in Studies on the Period, Amplitude and Type of the Uplift of the Qinghai-Xizang Plateau. Science Press, Beijing, pp. 118–130 (in Chinese).

Cavender, T., Coburn, M.M., 1992. Phylogenetic relationships of North American Cyprinidae. In: Mayden, R. (Ed.), Systematics, Historical Ecology, and American Freshwater Fishes. Stanford University Press, Stanford, CA, pp. 93–327.

Chakraborty, A., Iwatsuki, Y., 2006. Genetic variation at the mitochondrial 16S rRNA gene among *Trichiurus lepturus* (Teleostei: Trichiuridae) from various localities: preliminary evidence of a new species from West Coast of Africa. Hydrobiologia 563, 501–513.

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

Chen, Y.Y., 1998. Fauna Sinica, Osteichthys: Cypriniformes, Part II. Science Press, Beijing.

Cunha, C., Mesquita, N., Dowling, T.E., Gilles, A., Coelho, M.M., 2002. Phylogenetic relationships of Eurasian and American cyprinids using cytochrome *b* sequences. J. Fish Biol. 61, 929–944.

Dixon, M.T., Hillis, D.M., 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Mol. Biol. Evol. 10, 256–267.

Durand, J.D., Tsigenopoulos, C.S., Unlu, 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.

Felsenstein, J.P., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J.P., 1989. PHYLIP (phylogeny inference package). Version 3.2. Cladistics, 5, 164–166.

Galewski, T., Tilak, M., Sanchez, S., Chevret, P., Paradis, E., Douzery, E.J.P., 2006. The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. BMC Evol. Biol. 6, 80–96.

Galtier, N., 1996. Seaview and PHylo-win: two graphic tools for sequence alignment and molecular phylogeny. Comp. Appl. Bios. 12, 543–548.

Gilles, A., Lecointre, G., Faure, E., Chappaz, R., Brun, G., 1998. Mitochondrial phylogeny of European cyprinids: implications for their systematics, reticulate evolution, and colonization time. Mol. Phylogenet. Evol. 10, 132–143.

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.

Gosline, W.A., 1978. Unbranched dorsal-fin rays and subfamily classification of the fish family Cyprinidae. Occasional Papers the Museums of Zoology, University of Michigan, vol. 684, pp. 1–21.

Harris, P.H., Mayden, R.L., 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.

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

He, D., Chen, Y., Chen, Y., Chen, Z., 2004b. 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, D., Chen, Y., 2006. Biogeography and molecular phylogeny of the genus Schizothorax (Teleostei: Cyprinidae) in China inferred from cytochrome *b* sequences. J. Biogeogr. 33, 1448–1460.

Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47, 3–8.

Hillis, D.M., Huelsenbeck, J.P., 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. 83, 189–195.

Higgs, P.G., 1998. Compensatory neutral mutations and the evolution of RNA. Genetica 102, 91–101.

Higgs, P.G., 2000. RNA secondary structure: physical and computational aspects. Quart. Rev. Biophys. 33, 199–253.

Howes, G.J., 1991. Systematics and biogeography: an overview. In: WindWeld, I.J. (Ed.), Cyprinid Fishes. Systematics, Biology and Exploitation. Chapman & Hall, London, pp. 1–33.

Hudelot, C., Gowri-Shankar, V., Jow, H., Rattray, M., Higgs, P.G., 2003. RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. Mol. Phylogenet. Evol. 28, 241–252.

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

Jow, H., Hudelot, C., Rattray, M., Higgs, P., 2002. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Mol. Biol. Evol. 19, 1591–1601.

Kjer, K.M., 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53, 506–514.

Li, J., Wang, X., He, S., 2005. Phylogenetic studies of Chinese labeonine fishes (Teleostei: Cyprinidae) based on the mitochondrial 16S rRNA gene. Prog. Nat. Sci. 3, 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.

Liu, H., 2004. Phylogenetic relationships of the Cypriniformes tested by mtDNA 12S rRNA sequence variations. Acta Genetica Sin. 31, 137–142.

Maddison, W.P., Maddison, D.R., 1992. MacClade: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, Mass.

Mayden, R.L., Conway, K.W., Tang, K., Freyhof, J., Chamberlain, S., Haskins, M., Schneider, L., Sudkamp, M., Wood, R.M., Agnew, M., Bufalino, A., Sulaiman, Z., Miya, M., Saitoh, K., He, S., 2007. Phylogenetic relationships of *Danio* within the order Cypriniformes: a framework for comparative and evolutionary studies of a model species. J. Exp. Zool. B Mol. Dev. Evol.

Misof, B., Anderson, C.L., Buckley, T.R., Erpenbeck, D., Rickert, A., Misof, K., 2002. An empirical analysis of mt 16S rRNA covarion-like evolution in insects: site specific rate variation is clustered and frequently detected. J. Mol. Evol. 55, 460–469.

Moyer, G.R., Burr, B.M., Krajewski, C., 2004. Phylogenetic relationships of thorny catfishes (Siluriformes: Doradidae) inferred from molecular and morphological data. Zool. J. Linn. Soc-Lond. 140, 551–575.

Nelson, J.S., 2006. Fishes of The World, 4th ed. John Wiley & Sons, New York.

Nickrent, D.L., Parkinson, C.L., Palmer, J.D., Duff, R.J., 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol. Biol. Evol. 17, 1885–1895.

Ortí, G., Petry, P., Porto, J.A., Jegú, M., Meyer, A., 1996. Patterns of nucleotide change in mitochondrial ribosomal RNA genes and the phylogeny of piranhas. J. Mol. Evol. 42, 169–182.

Ortí, G., Meyer, A., 1997. The radiation of Characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. Syst. Biol. 46, 75–100.

Philippe, H., 2000. Opinion: long branch attraction and protest phylogeny. Protist 151, 307–316.

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

Rainboth, W.J., 1991. Cyprinids of South East Asia. In: Winfield, I.J., Nelson, J.S. (Eds.), Cyprinid Fishes. Systematics, Biology and Exploitation. Chapman & Hall, London, pp. 156–210.

Reid, G.M., 1982. The form, function and phylogenetic significance of the vomero-palatine organ in cyprinid fishes. J. Nat. Hist. 16, 497–510.

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

Rosenberg, M.S., Kumar, S., 2003. Taxon sampling, bioinformatics, and phylogenomics. Syst. Biol. 52, 119–124.

Rzhetsky, A., 1995. Estimating substitution rates in ribosomal RNA genes. Genetics 141, 771–783.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Clong: A Laboratory Manual, second ed. Cold Spring Harber Laboratory Press, New York.

Savill, N.J., Hoyle, D.C., Higgs, P.G., 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. Genetics 157, 399–411.

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

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

Simons, A.M., Mayden, R.L., 1998. Phylogenetic relationships of the western North American Phoxinins (Actinopterygii: Cyprinidae) as inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Mol. Phylogenet. Evol. 9, 308–329.

Simons, A.M., Berendzen, P.B., Mayden, R.L., 2003. Molecular systematics of North American phoxinin genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Zool. J. Linnean Soc. 139, 63–80.

Sloss, B.L., Billington, N., Burr, B.M., 2004. A molecular phylogeny of the Percidae (Teleostei: Perciformes) based on mitochondrial DNA sequence. Mol. Phylogent. Evol. 32, 545–562.

Springer, M.S., Hollar, L.J., Burk, A., 1995. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. Mol. Biol. Evol. 12, 1138–1150.

Springer, M.S., Douzery, E., 1996. Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. J. Mol. Evol. 43, 357–373.

Springer, M.S., Amrine, H.M., Burk, A., Stanhope, M.J., 1999. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition. Syst. Biol. 48, 65–75.

Susko, E., Field, C., Blouin, C., Roger, A.J., 2003. Estimation of rates-across-sites distributions in phylogenetic substitution models. Syst. Biol. 52, 594–603.

Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogeny inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.). Molecular Systematics, second ed., Sinauer Associates, Sunderland, MA, pp. 407–425.

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.

Telford, M.J., Wise, M.J., Gowri-Shankar, V., 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the Bilateria. Mol. Biol. Evol. 22, 1129–1136.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37, 221–244.

Tsigenopoulos, C.S., Karakousis, Y., Berrebi, P., 1999. The North Mediterranean Barbus lineage: phylogenetic hypotheses and taxonomic implications based on allozyme data. J. Fish Biol. 54, 267–286.

Tsigenopoulos, C.S., Berrebi, P., 2000. Molecular phylogeny of North Mediterranean freshwater barbs (genus Barbus: Cyprinidae) inferred from cytochrome *b* sequences: biogeographic and systematic implications. Mol. Phylogenet. Evol. 14, 165–179.

Van de Peer, Y., 1993. Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. J. Mol. Evol. 37, 221–232.

Vawter, L., Brown, W.M., 1993. Rates and pattern of base change in the small subunit ribosomal RNA gene. Genetics 134, 597–608.

Wang, H., Lee, S., 2002. Secondary structure of mitochondrial 12S rRNA among fish and its phylogenetic applications. Mol. Biol. Evol. 19, 138–148.

Wang, X., He, S., Chen, Y., 2002. Sequence variations of the S7 ribosomal protein gene in primitive cyprinid fishes: implication on phylogenetic analysis. Chinese Sci. Bull. 47, 1638–1643.

Wang, X., Li, J., He, S., 2007. Molecular evidence for the monophyly of East Asian of Cyprinidae (Teleostei: Cypriniformes) derived from the nuclear recombination activating gene 2 sequences. Mol. Phylogenet. Evol. 42, 157–170.

Waters, J.M., Lopez, J.A., Wallis, G.P., 2000. Molecular phylogenetics and biogeography of galaxiid fishes (Osteichthyes: Galaxiidae): dispersal, vicariance and the position of Lepidogalaxias salamandroides. Syst. Biol. 49, 777–795.

Whitfield, J.B., Cameron, S.A., 1998. Hierarchical analysis of variation in the mitochondrial 16S rRNA gene among Hymenoptera. Mol. Biol. Evol. 15, 1728–1743.

Wheeler, W.C., Honeycutt, R.L., 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Mol. Biol. Evol. 5, 90–96.

Wuyts, J., Perriere, G., Van de Peer, Y., 2004. The European ribosomal RNA database. Nucleic Acids Res. 32, 101–103.

Xia, X., Xie, Z., Kjer, K.M., 2003. 18S Ribosomal RNA and tetrapod phylogeny. Syst. Biol. 52, 283–295.

Xiao, H., Chen, S.Y., Liu, Z.M., Zhang, R.D., Li, W.X., Zan, R.G., Zhang, Y.P., 2005. Molecular phylogeny of Sinocyclocheilus (Cypriniformes: Cyprinidae) inferred from mitochondrial. DNA sequences. Mol. Phylogenet. Evol. 36, 67–77.

Yang, Z., 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. 39, 105–111.

Yang, Z., Kumar, S., 1996. Approximate methods for estimating the pattern of nucleotide substitution and variation of substitution rate among sites. Mol. Biol. Evol. 13, 650–659.

Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11, 367–372.
Yu, X., Zhou, T., Li, K., Li, Y., Zhou, M., 1987. On the karyosystematics of cyprinid fishes and a summary of fish chromosome studies in China. Genetica 72, 225–236.
Zardoya, R., Doadrio, I., 1998. Phylogenetic relationships of Iberian cyprinids: systematic and biogeographical implications. Proc. R. Soc. Lond. Ser. B 265, 1365–1372.
Zardoya, R., Doadrio, I., 1999. Molecular evidence on the evolutionary patterns of European Cyprinids. J. Mol. Evol. 49, 227–237.