SCIENCE CHINA
Life Sciences

• RESEARCH PAPER •

September 2012 Vol.55 No.9: 761–773
doi: 10.1007/s11427-012-4366-z

Cyprinid phylogeny based on Bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics

WANG XuZhen[1], GAN XiaoNi[1], LI JunBing[1], MAYDEN Richard L.[2] & HE ShunPing[1*]

[1]*Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China;*
[2]*Department of Biology, Saint Louis University, Saint Louis, Missouri 63103-2010, USA*

Received April 25, 2012; accepted July 16, 2012

Cyprinidae is the biggest family of freshwater fish, but the phylogenetic relationships among its higher-level taxa are not yet fully resolved. In this study, we used the nuclear recombination activating gene 2 and the mitochondrial 16S ribosomal RNA and cytochrome b genes to reconstruct cyprinid phylogeny. Our aims were to (i) demonstrate the effects of partitioned phylogenetic analyses on phylogeny reconstruction of cyprinid fishes; (ii) provide new insights into the phylogeny of cyprinids. Our study indicated that unpartitioned strategy was optimal for our analyses; partitioned analyses did not provide better-resolved or -supported estimates of cyprinid phylogeny. Bayesian analyses support the following relationships among the major monophyletic groups within Cyprinidae: (Cyprininae, Labeoninae), ((Acheilognathinae, ((Leuciscinae, Tincinae), Gobioninae)), Xenocyprininae). The placement of Danioninae was poorly resolved. Estimates of divergence dates within the family showed that radiation of the major cyprinid groups occurred during the Late Oligocene through the Late Miocene. Our phylogenetic analyses improved our understanding of the evolutionary history of this important fish family.

Cyprinidae, phylogeny, molecular dating, partitioned analyses

Citation: Wang X Z, Gan X N, Li J B, *et al*. Cyprinid phylogeny based on Bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. Sci China Life Sci, 2012, 55: 761−773, doi: 10.1007/s11427-012-4366-z

The family Cyprinidae is the largest freshwater fish family and includes an estimated 2420 species in about 220 genera [1]. The large number of species, wide geographic distribution, and considerable morphological diversity make the cyprinid fishes taxonomically difficult [2] and a challenge for cladistic analysis. The history of Cyprinidae classification was well documented by Hensel [3], and numerous efforts have been made to partition cyprinids into subfamilies using morphological or anatomical characteristics [2,4−8]. However, the systematic relationships among many cyprinid subfamilies are poorly understood, because the subfamilies are vaguely defined or supported by few morphological characteristics [2].

Cyprinidae has been conventionally divided into two major lineages, the cyprinine (barbine) and the leuciscine groups. Overall, morphology has provided few insights into cyprinid relationships below the family level and failed to reach agreement on the number and the monophyly of subfamilies within Cyprinidae. Chen *et al*. [8] published the cladistic evaluation of cyprinid subfamily relationships and the additional morphological studies by Cavender and Coburn [9] and Howes [2] attempted a coherent classification of all cyprinid groups, including the monotypic genus *Tinca*. In these previous studies, conflicting arrangements of the

*Corresponding author (email: clad@ihb.ac.cn)

subfamilies Tincinae, Rasborinae, and Gobioninae were proposed.

Recently, molecular phylogenetic analyses have been performed on Cyprinidae. In general, most of the molecular studies of European cyprinids [10–16] have been phylogenetically congruent. For example, all of these studies supported the nesting of Alburninae [2] within a paraphyletic Leuciscinae, but not the usual dichotomy between barbelled cyprinines (subfamilies Cyprininae, Gobioninae, and Rasborinae) and leuciscines lacking or sporadically possessing barbels (subfamilies Acheilognathinae, Leuciscinae, Cultrinae, and Alburninae) [2]. Because cyprinids are most diverse in Asiatic waters [17], phylogenetic studies that include Asian species would greatly advance cyprinid systematics [18–20]. Cunha *et al*. [18] identified an Asian group consisting of cultrins+acheilognathins+gobionins+xenocyprinins within the Cyprinidae using cytochrome b (Cytb) gene sequences. Other molecular phylogenies of East Asian cyprinids indicated two principal lineages within Cyprinidae and provided phylogenetic evidence for the monophyly of cultrins-xenocyprins and affiliated groups [19,20]. However, these molecular analyses were heavily based on partial mtDNA sequences, and resulted in phylogenetic trees with limited resolution and little discrimination among alternative phylogenetic hypotheses.

The current cyprinid classification developed in the absence of a strong phylogenetic framework and is largely morphology based; few revisions have resulted from recent molecular evidence, due to the limited taxon sampling in those studies. Some critical areas of Cyprinidae phylogeny and systematics remain unresolved. First, a majority of designated cyprinid subfamilies have not been tested for monophyly with either molecular or morphological data, and molecular data [18–20] has failed to support the monophyly of many morphologically-defined subfamilies, e.g., Rasborinae [8] and Leuciscinae [2,4]. Second, previous analyses have not agreed on the phylogenetic positions of Rasborinae, Tincinae, and Acheilognathinae. In recent molecular phylogenies, relationships among these subgroups remained unclear, because corresponding nodes were generally not statistically supported. Third, although the leuciscine and cyprinine subdivisions of Cyprinidae are widely accepted, the higher-level taxonomic relationships within these clades remain unresolved.

Molecular phylogenetic analyses of East Asian cyprinid resulted in substantial disagreement on the classification of subfamilies compared with the traditional taxonomy [19,20]. Therefore, extensive sampling of Asian cyprinids would provide further insights into the phylogenetic systematics of this family. The present paper used extensive taxon sampling and concatenated sequence data for the nuclear recombination activating gene 2 (RAG2) and the mitochondrial 16S ribosomal RNA (16S rRNA) and Cytb genes to reconstruct the phylogeny of cyprinids.

To analyze DNA sequence data-sets with multiple genes, partitioned phylogenetic analyses have become increasingly popular in recent years. Partitioned phylogenetic analyses use separate nucleotide substitution models (and associated parameters) for subsets of the data, to better explore partition-specific models of evolution and to reduce systematic error, thus yielding more accurate phylogenies. Generally, partitioned phylogenetic analyses are undertaken in a Bayesian framework [21], but recently, mixed-model search methods using maximum likelihood (ML) have become available [22]. Furthermore, an appropriately-partitioned data-set should be well modeled but not over-partitioned, because the over-parameterization (including over-partitioning) could result in parameter nonidentifiability, increased variance, improper posterior distributions, and undue influence of the priors [23]. Alternatives to Bayes factors for phylogenetic model selection that use explicit parameterization penalties are now available for partitioned analyses [23].

We performed ML and Bayesian analyses of partitioned data to reconstruct the phylogeny of cyprinid fishes, and also used relaxed molecular clock approaches to estimate the dates of cladogenetic events within the family. Our main aims were (i) to demonstrate how partitioning concatenated data affected phylogenetic reconstruction of cyprinid fishes; (ii) to test the monophyly of the currently-recognized subfamilies within Cyprinidae; and (iii) to discuss the taxonomic implications of the recovered clades.

1 Materials and methods

1.1 Taxon sampling and total DNA isolation

Our samples include 103 cyprinid species representing all major morphological groups and all 12 subfamilies within Cyprinidae [4]. Outgroup taxa were selected based on the consensus that Cypriniformes is a monophyletic group [24,25]. Therefore, six cypriniform fishes outside Cyprinidae were included in our analyses (Catostomidae, Balitoridae, Cobitidae, and Gyrinocheilidae) (Table 1).

Field-collected fish muscle or fin tissues were fixed in 95% ethanol and kept at −20°C in the laboratory until DNA extraction. Total genomic DNA was isolated from muscle or fin tissues using the phenol/chloroform extraction procedure [26].

1.2 DNA sequences collection and alignment

The nuclear RAG2 gene and mitochondrial genes were amplified from total DNA extracts via polymerase chain reaction (PCR) using published and/or optimized primers [27–29]. Reaction mixtures contained approximately 100 ng of DNA template, 5 μL of 10× reaction buffer, 2 μL dNTPs (each 2.5 mmol L^{-1}), 2.0 U *Taq* polymerase, and 1 μL of each oligonucleotide primer (10 μmol L^{-1} each), in a final volume 50 μL. The PCR amplification profile included an

Table 1 Cyprinid ingroup and cypriniform outgroup taxa used in this study and GenBank accession numbers of the sequence data[a)]

Subfamily	Taxa	Sampling location	Accession No.		
			RAG2	16S rRNA	Cytb
Barbinae	*Acrossocheilus beijiangensis*	Rong'an, Guangxi Zhuang Auto. Region	DQ366967	DQ845869	–
	Acrossocheilus elongates	Rong'an, Guangxi Zhuang Auto. Region	DQ366979	GQ406254	–
	Acrossocheilus hemispinus	Rong'an, Guangxi Zhuang Auto. Region	DQ366986	DQ845867	GQ406312
	Balantiocheilos melanopterus	Aquarium, Wuhan	DQ366933	GQ406255	–
	Barbodes huangchuchieni	Mengla, Yunnan Prov.	DQ366952	GQ406256	–
	Barbodes vernayi	Mengla, Yunnan Prov.	DQ366987	DQ845870	GQ406313
	Barbonymus schwanenfeldii	Aquarium, Wuhan	DQ366961	DQ845906	AF180823*
	Barbus barbus	France	DQ366990	DQ845879	AB238965*
	Barbus sp.	Africa	DQ366980	DQ845860	AF180842*
	Hampala macrolepidota	Mengla, Yunnan Prov.	DQ366965	DQ845863	DQ464974*
	Onychostoma gerlachi	Jinghong, Yunnan Prov.	DQ366963	DQ845862	GQ406314
	Onychostoma leptura	Xilin, Guangxi Zhuang Auto. Region	DQ366955	GQ406257	–
	Onychostoma macrolepis	Taian, Shandong Prov.	DQ366942	GQ406258	–
	Onychostoma sima	Hejiang, Sichuan Prov.	DQ366991	DQ845861	–
	Percocypris pingi pingi	Hejiang, Sichuan Prov.	DQ366962	GQ406259	–
	Puntius semifasciolatus	Jinghong, Yunnan Prov.	DQ366951	GQ406260	AY856116*
	Puntius conchonius	Aquarium	GQ406253	DQ845880	AY004751*
	Puntius tetrazona varieties	Aquarium	DQ366938	EU287909	EU287909*
	Sikukia stejnegeri	Mengla, Yunnan Prov.	DQ366931	DQ845872	GQ406315
	Sinocyclocheilus tingi	Fuxian Lake, Yunnan Prov.	DQ366978	DQ845866	AY854701*
	Spinibarbus hollandi	Tunxi, Anhui Prov.	DQ366973	DQ845865	AY195629*
	Tor douronensis	Menglun, Yunnan Prov.	DQ366945	DQ845877	DQ464986*
	Tor qiaojiensis	Yingjiang, Yunnan Prov.	DQ366970	DQ845873	GQ406316
	Tor sinensis	Mengla, Yunnan Prov.	DQ366936	DQ845876	FJ211164*
Cyprininae	*Carassius auratus*	Wuhan, Hubei Prov.	DQ366941	AB006953	AB006953*
	Cyprinus carpio	Tian'e, Guangxi Zhuang Auto. Region	DQ366994	X61010*	X61010*
	Cyprinus multitaeniata	Guiping, Guangxi Zhuang Auto. Region	DQ366939	DQ845845	–
	Procypris rabaudi	Hejiang, Sichuan Prov.	DQ366969	DQ845846	GQ406317
Labeoninae	*Cirrhinus molitorella*	Tengxian, Guangxi Zhuang Auto. Region	DQ366959	DQ845883	AY463098*
	Crossocheilus latius	Tengchong, Yunan Prov.	DQ366982	DQ845882	–
	Crossocheilus reticulates	Menglun, Yunnan Prov.	DQ366937	GQ406261	–
	Discogobio bismargaritus	Liuzhou, Guangxi Zhuang Auto. Region	DQ366947	DQ845890	GQ406318
	Discogobio brachyphysallidos	Jinxiu, Guangxi Zhuang Auto. Region	DQ366958	DQ845901	GQ406319
	Discogobio laticeps	Tian'e, Guangxi Zhuang Auto. Region	DQ366949	DQ845889	GQ406320
	Epalzeorhynchos frenatus rar	Aquarium, Jinghong	DQ366943	DQ845905	GQ406321
	Garra kempi	Chayu, Tibet Auto. Region	DQ366968	DQ845885	–
	Garra orientalis	Ledong, Hainan Prov.	DQ366957	DQ845884	GQ406322
	Garra taeniata	Jinghong, Yunnan Prov.	DQ366953	GQ406262	–
	Henicorhynchus lineatus	Menglun, Yunnan Prov.	DQ366935	GQ406263	GQ406323
	Labeo yunnanensis	Mengla, Yunnan Prov.	DQ366948	DQ845881	GQ406324
	Lobocheilus melanotaenia	Menglun, Yunnan Prov.	DQ366940	DQ845902	DQ464990*
	Osteochilus salsburyi	Rong'an, Guangxi Zhuang Auto. Region	DQ366971	DQ845892	GQ406325
	Parasinilabeo assimilis	Rong'an, Guangxi Zhuang Auto. Region	DQ366992	DQ845887	GQ406326
	Pseudocrossocheilus bamaensis	Tian'e, Guangxi Zhuang Auto. Region	DQ366993	DQ845895	GQ406327
	Ptychidio jordani	Tian'e, Guangxi Zhuang Auto. Region	DQ366974	DQ845893	GQ406328
	Rectoris posehensis	Dou'an, Guangxi Zhuang Auto. Region	DQ366975	DQ845891	GQ406329
	Semilabeo notabilis	Jinxiu, Guangxi Zhuang Auto. Region	DQ366983	DQ845886	GQ406330
	Sinilabeo rendahli	Yidu, Hubei Prov.	DQ366932	GQ406264	–
Schizothoracinae	*Gymnocypris eckloni eckloni*	Huanghe, Qinghai Prov.	DQ366950	DQ845853	AY463522*
	G. przewalskii przewalskii	Qinghai Lake, Qinghai Prov.	DQ366954	DQ845851	AY463523*
	Gymnodiptychus dybowskii	Yili, Xinjiang Uygur Auto. Region	DQ366956	DQ845859	AY463513*
	Schizopygopsis younghusbandi younghusbandi	Bomi, Tibet Auto. Region	DQ366976	GQ406265	AY463501*
	Schizothorax dulongensis	Guyong, Yunnan Prov.	DQ366985	DQ845849	AY954284*
	Schizothorax meridionalis	Yingjiang, Yunnan Prov.	DQ366989	DQ845847	AY954287*

(*To be continued on the next page*)

(*Continued*)

Subfamily	Taxa	Sampling location	Accession No.		
			RAG2	16S rRNA	Cytb
Schizothoracinae	*Schizothorax molesworthi*	Chayu, Tibet Auto. Region	DQ366946	DQ845848	DQ126130*
	Schizothorax myzostomus	Guyong, Yunnan Prov.	DQ366960	DQ845850	GQ406331
	Schizothorax waltoni	Chayu, Tibet Auto. Region	DQ366981	GQ406266	AY463518*
Leuciscinae	*Cyprinella lutrensis*	GN531	DQ367019	GQ406267	AB070206*
	Leuciscus leuciscus	France	DQ367007	GQ406268	AY509823*
	Phoxinus phoxinus	Europe	DQ367022	GQ406269	Y10448*
	Phoxinus lagowskii	Hengren, Liaoning Prov.	DQ367035	GQ406270	AB162650*
	Rutilus rutilus	France	DQ367003	GQ406271	AF095610*
	Pimephales promelas	GN530	DQ367000	GQ406272	AF117203*
	Rhinichthys atratulus	GN529	DQ367018	GQ406273	AF452078*
	Elopichthys bambusa	Taoyuan, Hunan Prov.	DQ367016	GQ406274	GQ406332
	Ochetobius elongates	Taoyuan, Hunan Prov.	DQ367012	GQ406275	AF309506*
	Luciobrama macrocephalus	Tengxian, Guangxi Zhuang Auto. Region	DQ367013	GQ406276	–
	Ctenopharyngodon idella	Hengxian, Guangxi Zhuang Auto. Region	DQ366996	GQ406277	AF051860*
	Mylopharyngodon piceus	Taoyuan, Hunan Prov.	DQ367011	GQ406278	AF051870*
	Squaliobarbus curriculus	Wuhan, Hubei Prov	DQ367021	GQ406279	AF051877*
	Tinca tinca	Europe	DQ367029	GQ406280	Y10451*
Hypophthalmichthyinae	*Aristichthys nobilis*	Wuhan, Hubei Prov.	DQ367038	GQ406281	AF051855*
	Hypophthalmichthys molitrix	Chenxi, Hunan Prov.	DQ367002	GQ406282	AF051866*
Xenocyprinae	*Distoechodon tumirostris*	Wuhan, Hubei Prov.	DQ366998	GQ406283	AF336308*
	Pseudobrama simony	Taoyuan, Hunan Prov.	DQ367028	GQ406284	AF036194*
	Xenocypris argentea	Taoyuan, Hunan Prov.	DQ367024	GQ406285	AP009059*
Danioninae	*Danio apogon*	Mengla, Yunnan Prov.	DQ367039	GQ406286	–
	Danio rerio		U71094*	AC024175*	AC024175*
	Hemigrammocypris rasborella	Japan	DQ367008	GQ406287	AF375863*
	Nicholsicypris normalis	Diaoluoshan, Hainan Prov.	DQ367034	GQ406288	–
	Opsariichthys bidens	Taoyuan, Hunan Prov.	DQ367014	GQ406289	DQ367044*
	Raiamas guttatus	Mengla, Yunnan Prov.	DQ366966	GQ406290	AF051875*
	Tanichthys albonubes	Aquarium, Wuhan	DQ367023	GQ406291	EF151121*
	Zacco platypus	Jinxiu, Guangxi Zhuang Auto. Region	DQ367010	GQ406292	AY245048*
Cultrinae	*Culter alburnus*	Taoyuan, Hunan Prov.	DQ367004	GQ406293	AP009060*
	Cultrichthys erythropterus	Lingshan, Guangxi Zhuang Auto. Region	DQ367037	GQ406299	AF051859*
	Megalobrama amblycephala	Wuhan, Hubei Prov.	DQ367025	GQ406294	AF051867*
	Pseudohemiculter dispar	Rong'an, Guangxi Zhuang Auto. Region	DQ367001	GQ406296	–
	Pseudolaubuca sinensis	Taoyuan, Hunan Prov.	DQ367017	GQ406297	–
	Rasborinus lineatus	Hengxian, Guangxi Zhuang Auto. Region	DQ367036	GQ406298	–
	Sinibrama macrops	Rong'an, Guangxi Zhuang Auto. Region	DQ367006	GQ406295	–
	Toxabramis swinhonis	Bobai, Guangxi Zhuang Auto. Region	DQ367027	GQ406300	DQ464972*
Gobiobotinae	*Gobiobotia abbreviate*	Tian'e, Guangxi Zhuang Auto. Region	DQ367033	GQ406301	–
	Gobiobotia filifer	Wuhan, Hubei Prov.	DQ367032	GQ406302	AY953002*
Gobioninae	*Abbottina rivularis*	Nanchong, Sichuan Prov.	DQ366995	GQ406303	AF051856*
	Coreius heterodon	Wuhan, Hubei Prov.	DQ367005	GQ406304	AY953000*
	Gobio gobio	France	DQ367015	GQ406305	AY426592*
	Pseudogobio vaillanti	Tian'e, Guangxi Zhuang Auto. Region	DQ366999	GQ406306	AY953019*
	Pseudorasbora parva	Jinxiu, Guangxi Zhuang Auto. Region	DQ366997	GQ406307	AF051873*
	Sarcocheilichthys sinensis sinensis	Hejiang, Sichuan Prov.	DQ367026	GQ406308	AY952983*
	Saurogobio dabryi	Changyang, Hubei Prov.	DQ367020	GQ406309	AY245091*
Acheilognathinae	*Paracheilognathus meridianus*	Hengxian, Guangxi Zhuang Auto. Region	DQ367009	GQ406310	–
	Rhodeus sp	Xilin, Guangxi Zhuang Auto. Region	DQ367031	GQ406311	DQ026430*
Outgroup	*Micronemacheilus pulcher*	Rong'an, Gaungxi Zhuang Auto. Region	DQ367041	DQ845921	DQ105199*
	Myxocyprinus asiaticus	Wuhan, Hubei Prov.	DQ367043	DQ845896	AY986503*
	Paramisgurnus dabryanus	Rong'an, Guangxi Zhuang Auto. Region	DQ367040	DQ845922	AY625701*
	Misgurnus sp		AY804103*	AB242171*	AB242171*
	Pseudogastromyzon fangi	Hengxian, Guangxi Zhuang Auto. Region	DQ367042	DQ845920	DQ105221*
	Gyrinocheilus sp		AY804074*	AB242164*	AB242164*

a) The cyprinid subfamilies follow those proposed by Chen [4]. * indicates sequences downloaded from GenBank, and – indicates missing data.

initial denaturation step at 94°C for 3 min, followed by 35 cycles of 30 s at 94°C, annealing for 30 s at 45–56°C (depending on the gene amplified), extension for 90 s at 72°C, and a final extension for 8 min at 72°C. Amplified DNA was fractionated by electrophoresis through 0.8% low-melting agarose gels, recovered, and purified using BioStar Glassmilk DNA purification Kit (Biostar International, Toronto, ON, Canada) according to the manufacturer's instructions. Nucleotide sequences were determined using purified PCR product. We generated most of the sequences used in this study, and some sequences for the 16S rRNA and Cytb genes were obtained from GenBank (Table 1).

For the two protein-coding genes, RAG2 and Cytb, multiple sequence alignments were performed using CLUSTAL X [30]. For the 16S rRNA gene, sequences were initially aligned using CLUSTAL X, then manually aligned based on secondary structural elements and conserved motifs by comparing to existing models of 16S rRNA secondary structure for cyprinid fishes [31–33]. All data-sets analyzed for this study are available on request from the first author.

1.3 Data partitions and model selection

We performed partitioned analyses using different nucleotide substitution models and associated parameter for each data subsets. We evaluated ten distinct partitioning strategies ranging from unpartitioned to a maximum of eight partitions (Table 2). Each partitioning strategies were denoted with the letter P followed by the number of data partitions. The unpartitioned (P1) analyses applied a single model of sequence evolution to all the data. The eight-partition (P8) analyses included separated substitution models for the stems and loops of 16S rRNA gene and for each codon position of Cytb and of RAG2.

Model selection was undertaken using PAUP [34] and ModelTest 3.7 [35]. The Akaike information criterion (AIC) weighting [36] determined the best-fit nucleotide model for each data partition. The initial tree used in ModelTest was drawn arbitrarily from a set of equally-parsimonious trees obtained with the complete data [23]. Because MrBayes 3.1.2 [21] only allows models with one, two, or six substitution rates, the AIC-selected model was often impossible to implement. Consequently we were forced to choose between under- and over-parameterized models. In these situations, a feasible solution is to select the best over-parameterized model to avoid the possible negative consequences of under-parameterization, e.g., underestimated branch lengths and consequent long-branch attraction [23]. This recommendation has been verified by several simulation studies that found few costs associated with model over-parameterization, at least within the framework of the general time-reversible (GTR) family of models [37,38]. The model GTR+I+Γ was applied to all partitions in our Bayesian and ML phylogenetic analyses.

1.4 Phylogenetic analyses of the Cyprinidae

Bayesian phylogenetic analyses were performed using the software MrBayes 3.1.2 [21]. A Metropolis-coupled Markov chain Monte Carlo (MCMC) process was undertaken for each data partition running simultaneously with a cold chain and three incrementally heated chains. The default setting for the heating parameter (T=0.2) in our preliminary analyses resulted in no or infrequent state exchanges between chains. When an alternative temperature regime (T=0.02) was used, successful state exchanges between chains improved in proportion to 40%–80%.

MCMC analyses of each data partition were run for 2×10^7 generations, with sampling every 1000 generations. We employed two strategies to confirm stationarity. First, we plotted log-likelihood scores, tree lengths, and all model parameter values against generation number using Tracer v. 1.4 (http://tree.bio.ed.ac.uk/software/tracer/) to graphically evaluate "burn-in". Second, MCMC convergence was assessed graphically using the cumulative function of AWTY [39]. The cumulative function was used to analyze the posterior probability (*PP*) support values for each clade to verify that these values were stable across all post-burn-in generations within each analysis. The PPs should stabilize once the Markov chain reaches stationarity, and substantial deviation of *PP*s from equilibrium values over time would indi-

Table 2 Partitioning strategies used in the Bayesian and maximum likelihood (ML) phylogenetic analyses, with their HMLi and –ln*L* scores

Partition strategy	Partition identity	HMLi (Bayesian)	–ln*L* (ML)
P1	All data combined	73918.6	73819.135
P2	mtDNA; RAG2	73252.4	73241.452
P3	16S; Cytb; RAG2	72874.8	72985.832
P4	16S stems; 16S loops; Cytb; RAG2	72510.5	72648.3
P5a	16S; Cytbpos1; Cytbpos2; Cytbpos3; RAG2	72177.9	72061.933
P5b	16S; Cytb; RAG2pos1; RAG2pos2; RAG2pos3	72496.1	72628.226
P6a	16S stems; 16S loops; Cytbpos1; Cytbpos2; Cytbpos3; RAG2	71806.61	71717.441
P6b	16S stems; 16S loops; Cytb; RAG2pos1; RAG2pos2; RAG2pos3	72130.61	72291.174
P7	16S; Cytbpos1; Cytbpos2; Cytbpos3; RAG2pos1; RAG2pos2; RAG2pos3	71790.49	71704.833
P8	16S stems; 16S loops; Cytbpos1; Cytbpos2; Cytbpos3; RAG2pos1; RAG2pos2; RAG2pos3	71419.06	71360.693

cate a lack of chain convergence. Our diagnoses suggested that chain convergence generally occurred within the first two million generations of each analysis. We followed a conservative approach by discarding the first 10 million generations as burn-in and using the remaining 10 million generations (10000 sampled trees) in all subsequent analyses. The 50% majority-rule consensus trees were generated with mean branch-length estimates, *PP* values for each node, credible sets of trees, and parameter estimates.

Trees resulting from our partitioned analyses that explicitly accommodated among-partition rate variation (APRV) had greater harmonic mean log likelihoods (HMLi) than those from equivalent analyses that did not accommodate APRV. Therefore, we employed the "prset ratepr=variable" option in MrBayes in all partitioned analyses. In all MCMC runs, we assigned uniform priors to trees and parameters of models of sequence evolution, and an exponential prior to branch lengths.

Partitioned and unpartitioned ML analyses were performed in RAxML [22]. Following the recommendation of McGuire *et al*. [23], we performed two sets of analyses for each partitioning strategy. In the first set of analyses, we estimated the ML values, which were used in the ML strategy-selection procedure, of the P8 Bayesian topology under each partitioning strategy. In the second set of analyses, we searched for the ML topology with the highest likelihood during 200 runs on distinct starting trees, then used 500 bootstrap replicates to measure support for the recovered clades. We employed the GTRGAMMAI substitution model in both sets of analyses.

1.5 Evaluation of alternative partitioning strategies

Alternative partitioning strategies were evaluated using four different criteria [23]: standard Bayes factors [40,41], a modified AICc [23], the Bayes information criterion (BIC) [42], and a decision-theoretic (DT) approach [43–45]. The Bayes factor for any pair of partitioned models was the ratio of their marginal likelihoods. Marginal likelihoods are difficult to calculate, but can be approximated by the HMLi [46]. Using ln-transformed Bayes factors, we accepted Bayes factors greater than 10 (2ln Bayes factors>10) as strong support for the more partitioned model. Because the relationships of HMLi's under alternative partitioning strategies are similar to the relationships of ML values [23,47,48], we substituted the HMLis for ML values in the AICc, BIC, and DT tests of partition strategies under Bayesian framework. The partitioning strategy preferred by AICc, BIC, or DT had the minimum observed value. To estimate branch-length on a fixed-topology in MrBayes, the program's branch-swapping functionality was disabled and node-slider was enabled (by resetting props).

We compared the optimal partitioning strategies selected by Bayes factors, AICc, BIC, and DT tests in Bayesian analyses with those preferred by hierarchical LRT (hLRT), AICc, BIC, and DT in ML analyses. To better compare ML and Bayesian strategy-selection procedures, RAxML and Bayesian analyses employed only the GTR+I+Γ substitution model. All ML comparisons were based on likelihoods calculated for the eight-partition Bayesian consensus tree, whereas Bayesian model criteria were computed in the context of optimized trees for each partitioning strategy (except DT). To apply these partition selection criteria to our ML and Bayesian analyses, we calculated the number of parameters in each data partition following the recommendation of McGuire *et al*. [23].

1.6 Testing alternative cyprinid phylogenetic hypotheses

Bayesian hypothesis testing [49] was used to test whether alternative hypotheses of higher-level cyprinid relationships recovered in our partitioned Bayesian analyses could be rejected by the combined data. Because Bayesian analysis infers the distribution of trees proportional to their PPs, commonly used statistical methods to compare alternative topologies, such as the approximately unbiased test [50], are not plausible under the Bayesian framework. The 95% credible sets of tree (sampled at stationarity) was built by using the "sumt" command in MrBayes. All trees were imported into PAUP and filtered by the phylogenetic hypothesis of interest; that hypothesis could not be rejected statistically when one or more trees in the 95% credible set compatible with the hypothesis.

1.7 Molecular dating of cyprinids

Rate heterogeneity among lineages in the concatenated dataset was evaluated using LRTs comparing log likelihoods of both constrained and unconstrained trees. We used the GTR+I+Γ substitution model. A strict molecular clock was rejected (P<0.005, degrees of freedom=107). Therefore, the relaxed molecular clock model of Sanderson's nonparametric rate smoothing (NPRS) method [51] was used to estimate divergence dates.

The NPRS implemented in the program r8s [52] was used to produce ultrametric trees. Divergence date estimates were based on the topology resulting from the unpartitioned Bayesian analysis. Powell's algorithm for optimizing the objective function and the additive penalty function were used. The 95% confidence intervals for the estimated ages were determined using 100 bootstrap pseudoreplicates of the combined data matrix using SEQBOOT in PHYLIP 3.5c [53]. While keeping the tree fixed, nodal depth (hence age estimates) of each pseudoreplicate was estimated by ML with the preferred model of molecular evolution [51]. For each node, the mean age was calculated from 100 bootstrap ages.

To estimate divergence times, we applied multiple fossil calibration points including (i) the root node of Cyprinidae

was constrained to a maximum of 55.8 million years ago (Mya) because the oldest reliable known fossils of Cyprinidae are from the Eocene [54]; (ii) the split between *Tinca* and the modern leuciscins was constrained to be a maximum of 18.0 Mya, because *Tinca* was described from the Middle Miocene [55,56] and a prominent turnover of European freshwater fish faunas represented by the appearance of modern *Palaeoleucisus* sp. and *Palaeocarassius* sp. (=aff. *Abramis* sp. vel aff. *Alburnus* sp.) happened about 17–18 Mya (the late early Miocene) [57]; (iii) a minimum of 1.81 Mya was assigned to the node subtending silver (*Hypophthalmichthys molitrix*) and bighead (*Aristichthys nobilis*) carp, and to the node subtending grass carp (*Ctenopharyngodon idella*) [58]; (iv) a minimum age of 3 Mya was used to define the origin of *Pseudorasbora* [58]; (v) a fixed date of 13 Mya was used to define the lineage leading to modern European *Barbus barbus* according to the fossil records of *Barbus* [10].

2 Results

We generated 4257 aligned base pairs (bp) of DNA sequence data representing three genes, the nuclear RAG2 1287 aligned bp and the 16S rRNA 1830 bp and Cytb 1140 bp. Of those sites, 2209 were variable and 1797 were parsimony informative. In the 16S rRNA gene, 190 sites were variable and 106 parsimony informative. The Cytb gene had 160 variable and 140 parsimony informative sites. The remaining 290 variable sites, 270 of which were parsimony informative, occurred in the RAG2 gene.

2.1 Effects of alternative partitioning strategies

The HMLi and $-\ln L$ were used to evaluate partitioning strategies in the Bayesian and ML analyses, respectively. In the present study, adding partitions substantially improved the HMLi and $-\ln L$ scores (Table 2), suggesting that simpler partitioning strategies were poorer fits to the data than more complicated partitioning strategies. For example, partitioning Cytb and RAG2 by codon positions dramatically improved HMLi and $-\ln L$. Partitioning the 16S rRNA gene by stems and loops improved HMLi and $-\ln L$ by about 360 and 330 log-units respectively (P4 vs. P3). Comparing the strategies with the same numbers of partitions (P5a vs. P5b, P6a vs. P6b) indicated that partitioning Cytb alone by codon was better than partitioning only RAG2 by codon. The P8 strategy, which partitioned the rRNA gene by stems and loops and the coding genes by codon position proved best in both Bayesian and ML analyses.

Despite differences in model fit, tree topologies inferred by Bayesian and ML methods using the ten partition strategies were almost identical to one other; the differences involved alternative placements of weakly supported nodes (*PP*<0.90 and bootstrap support<70%). The most dramatic difference in topology occurred in the position of *Danio* within Cyprinidae; the Bayesian P7 analysis weakly supported a basal position for *Danio*, unlike the other analyses. Tree length estimates varied only slightly across partitioning strategies, and no notable differences in PPs (Bayesian analyses) or bootstrap supports (ML analyses) were found among strategies. The number of strongly-supported ingroup nodes (*PP* values⩾0.95) decreased between the unpartitioned (73 of 95 ingroup nodes with *PP* values⩾0.95) and the maximally partitioned (63 of 93). Our analyses suggested that highly-partitioned Bayesian analyses had relatively poor performance in recovering well-supported cyprinid nodes.

2.2 Selecting the optimal partitioning strategy

Two extreme partitioning strategies were selected by DT and Bayes factors (hLRT), AICc, and BIC. The DT selected the unpartitioned P1 strategy; in contrast, the other criteria preferred the most partitioned (P8) strategy. The more-partitioned strategies considered in this study did not provide better-resolved or -supported estimates of cyprinid phylogeny, because all strategies resulted in similar topologies and node support values. The partitioning strategy employed in this analysis was not as critical as expected. However, the phylogenetic analysis based on P1 (preferred by DT in both Bayesian and ML frameworks) required fewer parameters to be estimated, and we inferred that P1 strategy was optimal for our Bayesian and ML analyses. Although adding partitions improved likelihood scores, partitioning had little effect on topology or node support. Much of the improvement in likelihood scores obtained with more extensive partitioning was probably associated with substitution model and base frequency parameter estimates (nuisance parameters in this context) rather than with more critical topology and branch-length estimates [23].

2.3 Phylogeny of the Cyprinidae

Bayesian analyses of a combined molecular dataset resulted in informative phylogenetic estimates for cyprinids (Figure 1). The monophyly of Cyprinidae was strongly supported with a *PP* of 1.0 in all analyses. The unpartitioned Bayesian analysis resulted in a well-resolved and -supported cyprinid phylogeny, with 73 of 95 ingroup nodes receiving *PP* values⩾0.95 and two additional nodes with *PP* values of 0.90–0.95.

The unpartitioned Bayesian analysis strongly supported several important clades within Cyprinidae. However, the position of *Danio* at the base of the leuciscines was poorly supported (*PP*=0.61), and the genus was basal to the entire family in the P7 analyses. Within the cyprinine lineage (Clade I) (*sensu* Howes [2]), the monophyly of labeonine fishes (Clade B) and sister relationship between labeonine and non-labeonine cyprinine clades (except *Procypris*, Clade A) were both strongly supported (*PP*=1.0), whereas the relationships within non-labeonine cyprinine clade

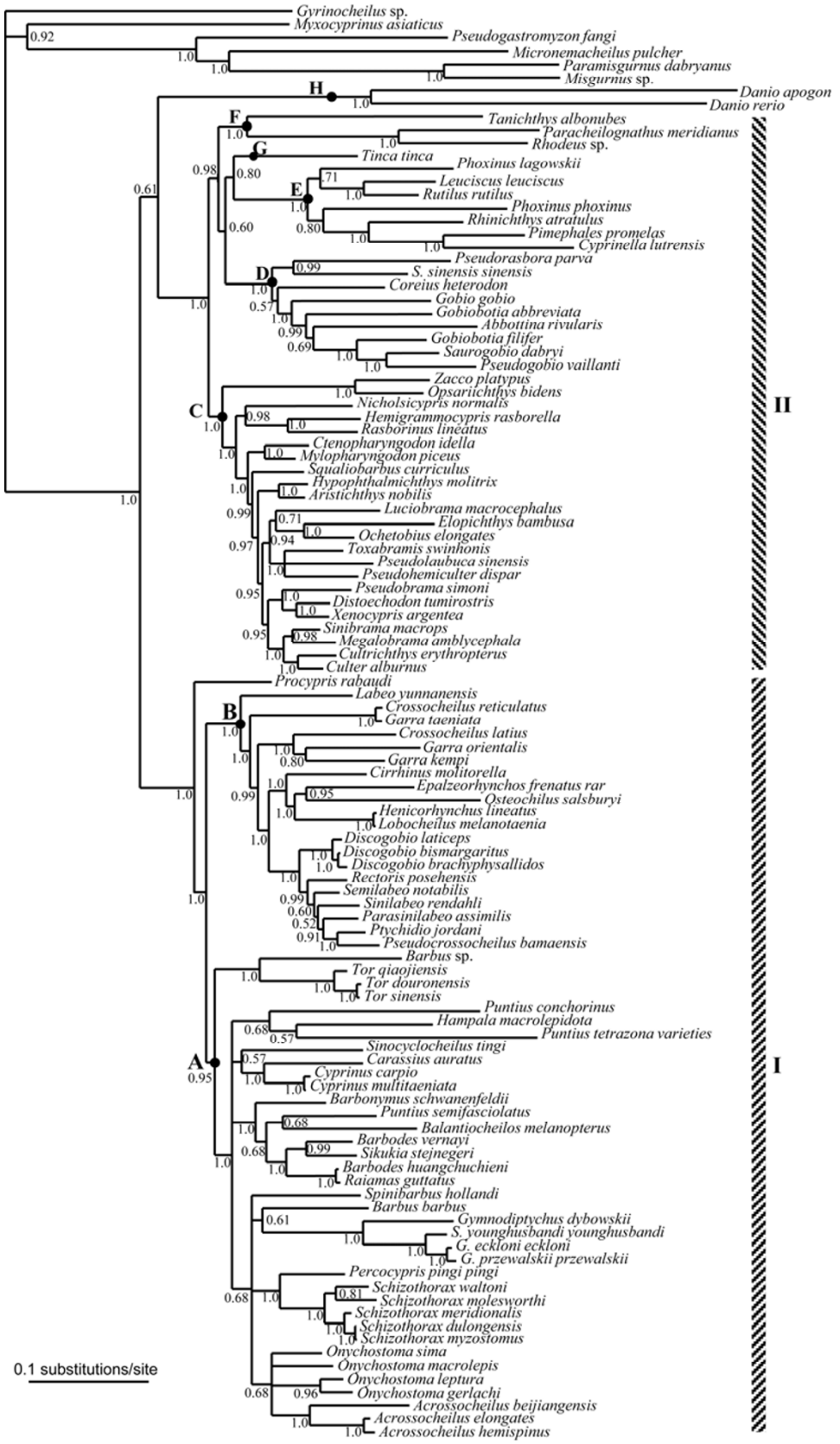

Figure 1 Phylogenetic tree of cyprinid fishes resulting from unpartitioned (P1) Bayesian analysis of three genes. Posterior probabilities values are shown at each node. Two recognized lineages within the family Cyprinidae are indicated by Roman numerals on the right side of the figure: (I) cyprinine lineage; (II) leuciscine lineage. Nodes for the recognized clades are marked with black dots and bold capital letters: A, cyprinine clade; B, labeonine clade; C, xenocyprinine clade; D, gobionine clade; E, leuciscine clade; F, acheilognathine (including *Tanichthys*) clade, G, tincine clade; H, danionine clade.

(Clade A) were less well supported, with several unresolved relationships.

Within the leuciscine lineage (Clade II), the cutrins were not supported as monophyletic. However, the East Asian endemic xenocyprinine taxa (including *Hypophthalmichthys*, *Aristichthys*, *Ctenopharyngodon*, *Mylopharyngodon*, *Ochetobius*, *Squaliobarbus*, *Elopichthys*, *Luciobrama*, *Culter*, *Cultrichthys*, *Sinibrama*, *Megalobrama*, *Pseudohemiculter*, *Toxabramis*, *Pseudolaubuca*, *Distoechodon*, *Xenocypris*, and *Pseudobrama*), formed a strongly-supported clade (Clade C, *PP*=1.0) in which the genera *Nicholsicypris*, *Rasborinus*, *Hemigrammocypris*, *Zacco*, and *Opsariichthys* were well resolved as the basal members. The North American and Eurasian leuciscins formed a strongly supported clade (Clade E, *PP*=1.0), but its sister relationship to the genus *Tinca* was weakly supported (*PP*=0.80). Clade D (*PP*=1.0), containing gobionins and gobiobotins, was weakly supported (*PP*=0.60) as sister to clade E+*Tinca*. Clade F, containing *Tanichthys*, *Rhodeus*, and *Paraacheilognathus*, was well supported (*PP*=1.0), as was the sister relationship between *Tanichthys* and the *Rhodeus*+*Paraacheilognathus* clade (*PP*=1.0). Furthermore, clade D, E, and F and *Tinca* formed a strongly supported clade (*PP*=0.98) sister to the strongly-supported East Asian endemic xenocyprinine clade (Clade C) (*PP*=1.0). The other nine partitioning strategies yielded Bayesian topologies largely in agreement with this tree.

The unpartitioned ML analysis (the strategy preferred by DT in an ML framework) and the more complex partitioning strategies resulted in phylogenetic trees highly similar to Figure 1, with the following exceptions: (i) the phylogenetic position of *Danio*; (ii) the deep branching pattern within Clade A; and (iii) support for the node subtending clade D, E, and F and *Tinca*. All of the important cyprinid clades recovered in Figure 1 were also well supported (bootstrap≥70%) in the ML analysis, except that the strongly-supported (*PP*=0.95) node for Clade A in the Bayesian analysis had relatively low bootstrap support (66%) in the ML analysis.

2.4 Divergence dates of cyprinid clades

Table 3 lists the divergence times (with 95% confidence intervals) estimated in r8s for the main nodes marked in Figure 2. As estimated by our combined data, Cyprinidae appeared in East Asia around 42.38 (43.13–41.64) Mya. The cyprinine and leuciscine lineages separated an estimate of 27.36 (27.84–26.89) Mya. Our dating results suggested that the cyprinine and leuciscine lineages began to diversify simultaneously (CYN node, 20.45 Mya; LEU node, 20.51 Mya). Within the cyprinine lineage, radiation of Labeoninae (node LAB) occurred 14.98 Mya, while the Cyprininae (node CYT) diverged earlier (18.46 Mya). Radiation of the clades Acheilognathinae (node ACH), Gobioninae (node GOB), Leuciscinae (node LES), and Xenocyprinine (node XEN) occurred from 18.80–12.02 Mya.

3 Discussion

3.1 Performance of alternative partitioning strategies

For the datasets composed of multiple genes and/or gene regions, partitioned phylogenetic analyses may greatly reduce mismodeling and systematic errors relative to analyses specifying a single model. Comparison of the 95% credible intervals of parameters sampled from the posterior distributions of the P1 and P8 analyses found significant heterogeneity, indicating that including more partitions greatly improved the Bayesian and ML likelihoods in this study. Numerous instances of non-overlap could be found in the credible intervals for different partitioning schemes. Based on the parameter estimates, partitioning the Cytb codon positions improved the HMLi and $-\ln L$ scores more substantially than partitioning the RAG2 or 16S rRNA genes.

Although partitioning substantially improved likelihood scores, its effect on topology and node support was minimal. In our Bayesian analyses, increased partitioning decreased the estimated PPs of some nodes relative to the P1 strategy. For example, seven of the ingroup nodes that had *PP* values

Table 3 Divergence time estimates and their 95% confidence interval for key nodes in the cyprinid phylogeny (Figure 2)

Clade	Node	Age estimates (Mya)	
		Mean±SD	95% interval
Cyprinidae, except outgroup	CYD	42.38±3.74	41.64–43.13
Cyprinine lineage	CYN	20.45±1.41	20.17–20.73
Cyprinine clade	CYT	18.46±1.04	18.25–18.67
Labeonine clade	LAB	14.98±0.95	14.79–15.17
Leuciscine lineage	LEU	20.51±1.22	20.27–20.76
Clades acheilognathine, gobionine, leuciscine, and the genus *Tinca*	NOR	19.41±0.88	19.23–19.58
Leuciscine clade	LES	12.02±0.63	11.90–12.15
Gobionine clade	GOB	14.13±0.88	13.95–14.30
Acheilognathine clade	ACH	17.11±0.97	16.92–17.30
Leuciscine clade and the genus *Tinca*	TIN	17.90±0.40	17.82–17.98
Xenocyprinine clade (East Asian endemic clade)	XEN	18.80±1.12	18.58–19.02

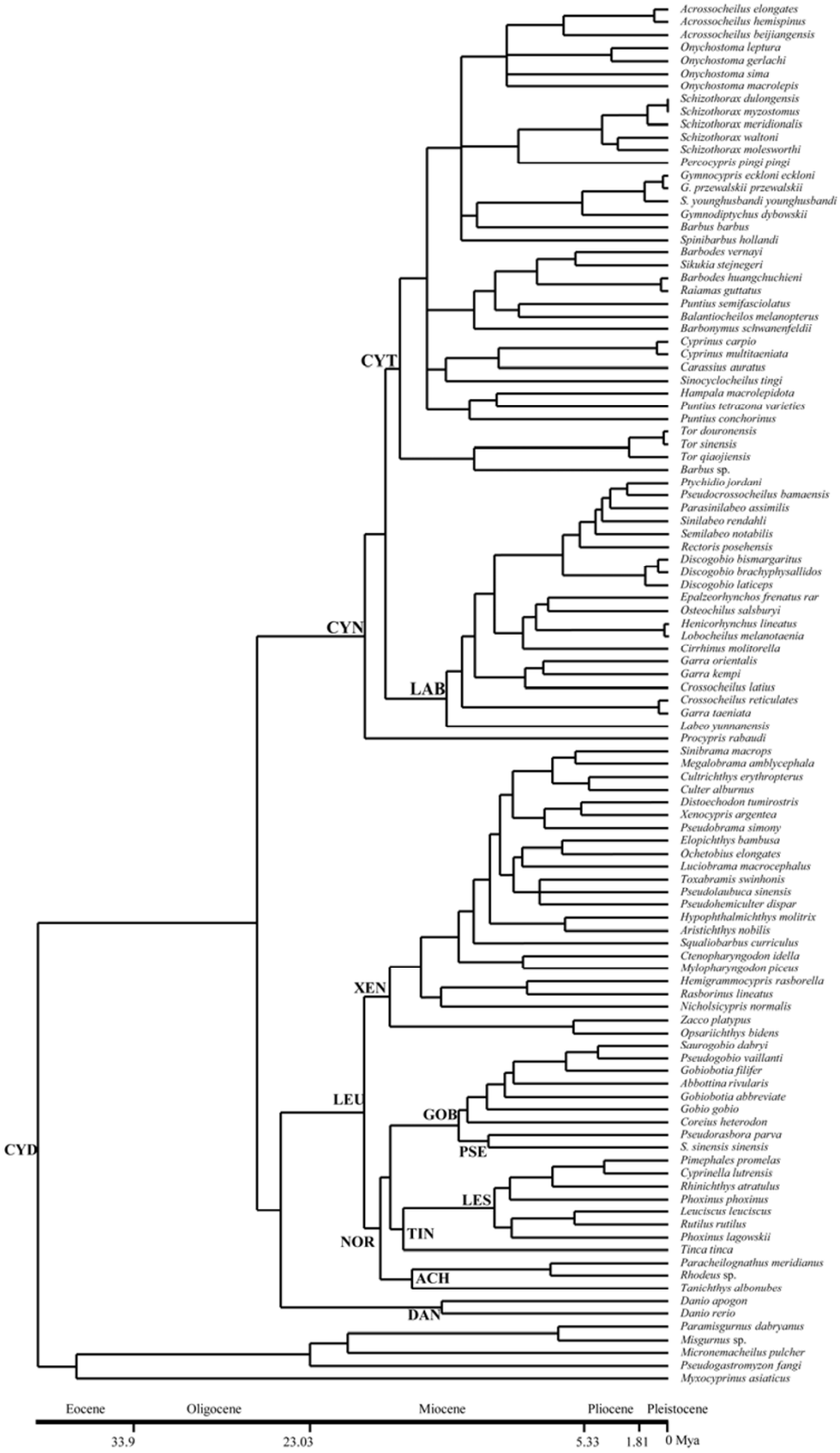

Figure 2 Phylogeny of cyprinid fishes with divergence time estimates. The chronogram is the tree from the unpartitioned Bayesian analysis with dates estimated using nonparametric rate smoothing in the program R8s. Node labels are defined in Table 3, where mean divergence dates and 95% confidence intervals for key nodes are listed.

of 1.0 in the P1 tree had lower support values under the most partitioned (P8) strategy. Phylogeneticists are concerned about appropriate partitioning in their analyses, because poor topology and confidence estimates can result from poorly- or overly-partitioned models. Although improved modeling could decrease the amount of systematic error under a given partitioning strategy, random error could significantly impact phylogeny and confidence estimates. The ideal partition size for optimal phylogenetic estimates is still unclear. For our cyprinid dataset, we concluded that most of the improvement in HMLi and $-\ln L$ estimates with greater partitioning was associated with better estimation of nuisance parameters, such as base frequencies and substitution rates.

We compared ten partitioning strategies in both Bayesian and ML frameworks, and four alternative model-selection criteria were employed to screen the best-fitting strategy. The standard Bayes factor/hLRT and AICc imposed relatively weak penalties for additional parameterization and consequently selected the most complex partitioning strategy, whereas the more stringent BIC and DT criteria preferred the most- and least-partitioned models, respectively. The DT method incorporates relative branch-length error as a performance measure. Therefore, under the DT framework, if a less-partitioned model returned nearly identical branch length estimates to those of a model with more partitions, there would be little difference in phylogenetic estimates between the models. The performance-based DT criterion selected the unpartitioned strategy in our analyses, indicating that there were probably no improvements in branch length estimates in our partitioned (P2–P8, Table 2) analyses compared with unpartitioned analyses.

3.2 Phylogenetic framework for systematics of the Cyprinidae

As expected, the monophyly of the family Cyprinidae was recovered with strong Bayesian *PP* and ML bootstrap support. Our phylogeny established a higher-level framework for Cyprinidae and revealed several well-supported groupings.

One large clade within Cyprinidae was the well-supported cyprinine lineage (Clade I, Figure 1). All taxa in this clade were members of the previously-recognized subfamilies, Barbinae, Cyprininae, Labeoninae, and Schizothoracinae [4]. Although the basal relationships within this clade have been contentious due to disagreement between molecular and morphological phylogenetic studies, our data consistently supported the monophyly of the cyprinine clade. Within the cyprinines, our analyses provided robust evidence for the monophyly of Labeoninae as the currently recognized. However, in all of our analyses, the cyprinine, barbine, and schizothoracine fishes (except *Procypris*) were nested within one clade (Clade A) sister to the labeonine clade. In another analysis with more cyprinine samples (unpublished data), two clades were strongly recovered: the Labeoninae and the Cyprininae, containing the barbins, cyprinins (including *Procypris*), and schizothoracins.

Another well-supported primary clade of Cyprinidae resolved in all analyses was the leuciscine lineage (Clade II, Figure 1). Within this clade, all of our analyses provided substantial resolution and support for the monophyly of Gobioninae (including *Gobiobotia*), Acheilognathinae, Leuciscinae, and Xenocyprininae, the latter is endemic to East Asia. Although Gobioninae, Acheilognathinae, and Leuciscinae were each strongly supported, the relationships among them were weakly resolved and differed among analyses. These three clades, together with *Tinca*, formed a clade sister to the Xenocyprininae. The placement of *Tinca* within Cyprinidae has proved to be taxonomically problematic in previous studies [2]. In contrast to studies based on morphological [8,9] and molecular [10–12,20] data, our analyses strongly supported a clade comprised of *Tinca*, leuciscini, Gobionini, and Acheilognathini, within which *Tinca* was weakly supported as sister to leuciscini.

Not surprisingly, the monophyly of the danionine (rasborine *sensu* Howes [2]) fishes was rejected by the present analyses. Morphologically, "Danioninae" contains a large assemblage of taxa, most of which cannot be accommodated by other subfamilies [2]. Furthermore, a recent molecular phylogeny indicated that Danioninae was not monophyletic; putative members were scattered throughout Cyprinidae [59]. Thus, we suggest that the East Asia endemics, such as *Zacco*, *Opsariichthys*, and *Nicholsicypris* should be excluded from a redefined Danioninae .

In the recent taxonomic revision of cyprinid (or cyprinoid) fishes by Chen and Mayden [60], the recognition of 10 families (including the Psilorhynchidae) was recommended. Of these groups, six (i.e., Acheilognathinae, Leuciscinae, Gobioninae, Cultrinae, Tincinae, and Rasborinae) were also supported in the present analyses (Figure 1). The Psilorhynchidae and Leptobarbidae were not included in our analyses, and the Cultrinae and Rasborinae referred to Xenocyprininae and Danioninae, respectively, in our study. Our data suggested that the Cyprinae recognized by Chen and Mayden [60] could be further divided into two clades, Cyprininae and Labeonine, and that the Tanichthyidae should be included in the Acheilognathinae. Unlike Chen and Mayden, we do not recommend that these groups be elevated from subfamily to family level, but prefer to retain these clades within Cyprinidae.

Previous morphological studies consistently supported two major lineages within Cyprinidae, i.e., barbeled cyprinines and (usually) non-barbeled leuciscines, although the subgroups included in each lineage and the relationships among subgroups have differed among studies [2,8,9]. However, recent molecular studies have disagreed with the morphological placement of Danionine (Rasborinae). The placement of Danioninae to the leuciscine clade was indicated in some prior morphological and molecular phylogeny

[8,9,60], and was weakly supported in our cyprinid phylogeny (Figure 1). Another study placed Danioninae within the cyprinine [2], while other molecular phylogenetic analyses placed it at the base of the cyprinids [12,13,61]. The disputed phylogenetic placement of Danioninae may be mainly due to different taxon sampling in these previous molecular phylogenies. Our data indicated that Danioninae represents a lineage within Cyprinidae, that is distinct from the well-accepted cyprinine and leuciscine lineages. A basal position for Danioninae within Cyprinidae (as recovered in the P7 Bayesian analysis) could not be rejected by Bayesian hypothesis testing of alternatives phylogenies generated from our combined data. A total of 6414 of 18710 trees in the 95% credible set were congruent with the hypothesis that Danioninae was basal within cyprinids.

3.3 Phylogenetic history of cyprinid clades

Based on the distribution of fossil cyprinids, an Eocene origin for cyprinids was proposed [54]. Consistent with this hypothesis, our molecular dating analyses also indicated that cyprinids originated in the Middle Eocene (around 42 Mya). Within the family, the cyprinine linage appeared in the early Late Oligocene (around 27 Mya) and the leuciscine lineage in the Late Oligocene (about 26–25 Mya).

The radiation of Labeoninae, the major cyprinine clade, occurred in the early Middle Miocene, and Cyprininae was estimated to have diversified in the late Early Miocene. Within the leuciscine lineage, the divergence between Xenocyprininae and the lineage comprising Leuciscinae, Tincinae, Gobioninae, and Acheilognathinae, occurred in the Early Miocene (about 20 Mya). According to our age estimates, the Acheilognathini, Gobionini, and Leuciscini radiated during the Middle Miocene (around 18–12 Mya).

We thank Dr. Wang JiangXin for his assistance in improving the presentation of our manuscript. This work was supported by the National Natural Science Foundation of China (Grant No. 30770300) and Chinese Academy of Sciences (Grant No. KSCX2-EW-Q-12), and the Cypriniformes Tree of Life Initiative supported by the USA National Science Foundation (Grant No. EF-0431326).

1 Nelson J S, ed. Fishes of the World. New York: John Wiley and Sons Inc., 2006
2 Howes G J. Systematics and biogeography: an overview. In: Winfield I J,Nelson J S, eds. Cyprinid Fishes: Systematics, Biology and Exploitation. London: Chapman and Hall, 1991. 1–33
3 Hensel K. Review of the classification and of the opinions on the evolution of Cyprinoidei (Eventognathi) with an annotated list of genera and subgenera described since 1921. Annot Zool Bot, 1970, 57: 1–45
4 Chen Y Y, ed. Fauna Sinica, Osteichthys: Cypriniformes (Part II). Beijing: Science Press, 1998
5 Gosline W A. Unbranched dorsal-fin rays and subfamily classification in the fish family Cyprinidae. Occas Pap Mus Zool Univ Mich, 1978, 684: 1–21
6 Chu Y T. Comparative studies on the scales and on the pharyngeals and their teeth in Chinese Cyprinids, with particular reference to taxonomy and evolution. Biol Bull St John's Univ (Shanghai), 1935, 2: 1–225
7 Wu X, ed. The Cyprinid Fishes of China (in Chinese). Shanghai: Shanghai Science and Technology Press, 1964
8 Chen X L, Yue P Q, Lin R D. Major groups within the family Cyprinidae and their phylogenetic relationships. Acta Zootaxon Sin, 1984, 9: 424–440
9 Cavender T M, Coburn M M. Phylogenetic relationships of North American Cyprinidae. In: Mayden R L, ed. Systematics, Historical Ecology and North American Freshwater Fishes. Stanford, California: Stanford University Press, 1992. 293–327
10 Zardoya R, Doadrio I. Molecular evidence on the evolutionary and biogeographical patterns of European Cyprinids. J Mol Evol, 1999, 49: 227–237
11 Briolay J, Galtier N, Brito R M, *et al.* Molecular phylogeny of Cyprinidae inferred from cytochrome b DNA sequences. Mol Phylogenet Evol, 1998, 9: 100–108
12 Gilles A, Lecointre G, Faure E, *et al.* Mitochondrial phylogeny of the European Cyprinids: implications for their systematics, reticulate evolution, and colonization time. Mol Phylogenet Evol, 1998, 10: 132–143
13 Gilles A, Lecointre G, Miquelis A, *et al.* Partial combination applied to phylogeny of European Cyprinids using the mitochondrial control region. Mol Phylogenet Evol, 2001, 19: 22–33
14 Zardoya R, Doadrio I. Phylogenetic relationships of Iberian Cyprinids: systematic and biogeographical implications. Proc R Soc B Biol Sci, 1998, 265: 1365–1372
15 Durand J D, Tsigenopoulos C S, Unlu E, *et al.* Phylogeny and biogeography of the family Cyprinidae in the Middle East inferred from cytochrome b DNA- evolutionary significance of this region. Mol Phylogenet Evol, 2002, 22: 91–100
16 Hanfling B, Brandl R. Phylogenetics of European Cyprinids: insights from allozymes. J Fish Biol, 2000, 57: 265–276
17 Fu C Z, Wu J H, Chen J K, *et al.* Freshwater fish biodiversity in the Yangtze River Basin of China: patterns, threats and conservation. Biodivers Conserv, 2003, 12: 1649–1685
18 Cunha C, Mesquita N, Dowling T E, *et al.* Phylogenetic relationships of Eurasian and American Cyprinids using cytochrome b sequences. J Fish Biol, 2002, 61: 929–944
19 He S, Liu H, Chen Y, *et al.* Molecular phylogenetic relationships of Eastern Asian Cyprinidae (Pisces: Cypriniformes) inferred from cytochrome b sequences. Sci China Ser C-Life Sci, 2004, 47: 130–138
20 Liu H, Chen Y. Phylogeny of the East Asian Cyprinids inferred from sequences of the mitochondrial DNA control region. Can J Zool, 2003, 81: 1938–1946
21 Ronquist F, Huelsenbeck J P. Mrbayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 2003, 19: 1572–1574
22 Stamatakis A. Raxml-Vi-Hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics, 2006, 22: 2688–2690
23 McGuire J A, Witt C C, Altshuler D L, *et al.* Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. Syst Biol, 2007, 56: 837–856
24 Fink S V, Fink W L. Interrelationships of the ostariophysan fishes (Teleostei). Zool J Linn Soc, 1981, 72: 297–353
25 Wang X, Wang J, He S, *et al.* The complete mitochondrial genome of the Chinese hook snout carp *Opsariichthys bidens* (Actinopterygii: Cypriniformes) and an alternative pattern of mitogenomic evolution in vertebrate. Gene, 2007, 399: 11–19
26 Sambrook J, Fritsch E, Maniatis T, ed. Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press, 1989
27 Xiao W H, Zhang Y P, Liu H Z. Molecular systematics of Xenocyprinae (Teleostei : Cyprinidae): taxonomy, biogeography, and coevolution of a special group restricted in East Asia. Mol Phylogenet Evol, 2001, 18: 163–173
28 Li J, Wang X, Kong X, *et al.* Variation patterns of the mitochondrial

16s rRNA gene with secondary structure constraints and their application to phylogeny of Cyprinine fishes (Teleostei: Cypriniformes). Mol Phylogenet Evol, 2008, 47: 472–487
29 Lovejoy N R, Collette B B. Phylogenetic relationships of new world needlefishes (Teleostei: Belonidae) and the biogeography of transitions between marine and freshwater habitats. Copeia, 2001, 2: 324–338
30 Thompson J D, Gibson T J, Plewniak F, *et al.* The Clustal_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res, 1997, 25: 4876–4882
31 De Rijk P, Wuyts J, Van de Peer Y, *et al.* The European large subunit ribosomal RNA database. Nucleic Acids Res, 2000, 28: 177–178
32 Gutell R R, Gray M W, Schnare M N. A compilation of large subunit (23s and 23s-like) ribosomal RNA structures. Nucleic Acids Res, 1993, 21: 3055–3074
33 Gutell R R, Fox G E. A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res, 1988, 16: r175–269
34 Swofford D L. Paup*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0b10. 2003
35 Posada D, Crandall K A. Modeltest: testing the model of DNA substitution. Bioinformatics, 1998, 14: 817–818
36 Posada D, Buckley T R. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol, 2004, 53: 793–808
37 Lemmon A R, Moriarty E C. The importance of proper model assumption in Bayesian phylogenetics. Syst Biol, 2004, 53: 265–277
38 Huelsenbeck J, Rannala B. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst Biol, 2004, 53: 904–913
39 Wilgenbusch J, Warren D, Swofford D. Awty: A System for Graphical Exploration of Mcmc Convergence in Bayesian Phylogenetic Inference. 2004
40 Kass R E, Raftery A E. Bayes factors. J Am Stat Assoc, 1995, 90: 773–795
41 Nylander J A A, Ronquist F, Huelsenbeck J P, *et al.* Bayesian phylogenetic analysis of combined data. Syst Biol, 2004, 53: 47–67
42 Schwarz G. Estimating the dimension of a model. Ann Stat, 1978, 6: 461–464
43 Minin V, Abdo Z, Joyce P, *et al.* Performance-based selection of likelihood models for phylogeny estimation. Syst Biol, 2003, 52: 674–683
44 Abdo Z, Minin V N, Joyce P, *et al.* Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. Mol Biol Evol, 2005, 22: 691–703
45 Sullivan J, Abdo Z, Joyce P, *et al.* Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. Mol Biol Evol, 2005, 22: 1386–1392
46 Raftery A. Hypothesis testing and model selection. In: Gilks W R, Spiegelhalter D J, Richardson S, eds. Markov Chain Monte Carlo in Practice. London: Chapman and Hall, 1996. 163–187
47 Castoe T A, Sasa M M, Parkinson C L. Modeling nucleotide evolution at the mesoscale: the phylogeny of the neotropical pitvipers of the *Porthidium* group (Viperidae: Crotalinae). Mol Phylogenet Evol, 2005, 37: 881–898
48 Castoe T A, Parkinson C L. Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). Mol Phylogenet Evol, 2006, 39: 91–110
49 Brandley M C, Schmitz A, Reeder T W. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst Biol, 2005, 54: 373–390
50 Shimodaira H. An approximately unbiased test of phylogenetic tree selection. Syst Biol, 2002, 51: 492–508
51 Sanderson M J. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol, 1997, 14: 1218–1231
52 Sanderson M J. R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics, 2003, 19: 301–302
53 Felsenstein J. Phylip (Phylogeny Inference Package) Version 3.5 C. 1993
54 Cavender T M. The fossil record of the Cyprinidae. In: Winfield I J, Nelson J S, eds. Cyprinid Fishes: Systematics, Biology and Exploitation. London: Chapman and Hall, 1991. 34–54
55 Hierholzer E, Mörs T. Cypriniden-Schlundzähne (Osteichthyes: Teleostei) aus dem Tertiär von Hambach (Niederrheinische Bucht, Nw-Deutschland). Palaeontographica, Abteilung A, 2003, 269: 1–38
56 Sytchevskaya E. Freshwater Ichthyofauna of the Neogene of Mongolia. Tr Sovm Sovets-Mongol Paleontol Eksped, 1989, 39: 1–144
57 Schulz-Mirbach T, Reichenbacher B. Reconstruction of oligocene and neogene freshwater fish faunas—an actualistic study on cypriniform otoliths. Acta Palaeontol Pol, 2006, 51: 283–304
58 Liu H, Su T. Pliocene fishes from the Yushe Basin, Shanxi. Vertebr Palasiat, 1962, 6: 1–25
59 Tang K L, Agnew M K, Hirt M V, *et al.* Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). Mol Phylogenet Evol, 2010, 57: 189–214
60 Chen W J, Mayden R L. Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the world's largest clade of freshwater fishes: further evidence from six nuclear genes. Mol Phylogenet Evol, 2009, 52: 544–549
61 Wang X, Li J, He S. Molecular evidence for the monophyly of East Asian groups of Cyprinidae (Teleostei: Cypriniformes) derived from the nuclear recombination activating gene 2 sequences. Mol Phylogenet Evol, 2007, 42: 157–170