Molecular Phylogenetics and Evolution
Vol. 19, No. 1, April, pp. 22–33, 2001
doi:10.1006/mpev.2000.0916, available online at http://www.idealibrary.com on IDEAL®

Partial Combination Applied to Phylogeny of European Cyprinids Using the Mitochondrial Control Region

André Gilles,* Guillaume Lecointre,† Anne Miquelis,* Matthias Loerstcher,* Rémi Chappaz,* and Guy Brun*

*Laboratoire d'Hydrobiologie, Université de Provence, 1 Place Victor Hugo, 13331 Marseille, France; and †Laboratoire d'Ichtyologie Générale et Appliquée et Service de Systématique Moléculaire (I.F.R. C.N.R.S. 1541), Muséum National d'Histoire Naturelle, 43 Rue Cuvier, 75231 Paris Cedex 05, France

Received February 21, 2000; revised November 13, 2000

Previous molecular phylogenies of European cyprinids led to some solid facts and some uncertainties. This study is based on a stretch of more than 1 kb in the mitochondrial control region newly sequenced for 35 European cyprinids and on previous cytochrome *b* and 16S rDNA data. The trees based on the control region are more accurate and robust than those obtained from the two other genes. Character incongruence among the three genes was tested using the incongruence length difference (ILD) test. Iterative removals of individual sequences followed by new ILD tests identified two sequences responsible for statistically significant incongruence. A partial combination was conducted, that is, a combination of the three data sets, removing the two sequences previously identified. The phylogenetic analysis of this partial combination gives a more robust and resolved picture of subfamilial interrelationships. The Rasborinae are the sister group of all other cyprinids. The monophyletic Cyprininae emerges next. *Tinca tinca* first and then *Rhodeus* are the sister groups of all the remaining nonrasborine and noncyprinine species. *Gobio* is the sister group of the Leuciscinae, in which the Phoxinini are the sister group of the Leuciscini. Within the Leuciscini, the genus *Leuciscus* and the subfamily Alburninae are both paraphyletic. The Rasborinae are the most basal cyprinid subfamily and the Tincinae are not the sister group of the Cyprininae. These two results challenge only two anatomical characters, which need to be reinterpreted or considered as homoplastic in cyprinid evolution: the modification of the first pleural rib and its parapophysis and the bony composition of the interorbital septum.

Key Words: **Cyprinidae; control region; d-loop; cytochrome *b*; 16S mtDNA; mitochondrial phylogeny.**

INTRODUCTION

Phylogenetic interrelationships of cyprinid subfamilies have been investigated from both morphological and molecular perspectives. Chen *et al.* (1984) first proposed hypotheses of cyprinid interrelationships based on morphological data and Cavender and Coburn (1992) performed a cladistic analysis of 47 morphological and anatomical characters leading to the most parsimonious tree shown in Fig. 1a. These authors also reanalyzed the matrix of Chen *et al.* (1984) and found the most parsimonious tree shown in Fig. 1b. The two morphological data sets differed only in the relationships of danionines (rasborines) and tincines. Previous molecular investigations resolved some, but not all groups. Studies based mostly on European cyprinids (Briolay *et al.*, 1998; Gilles *et al.*, 1998; Zardoya and Doadrio, 1999) found that (1) the Rasborinae are the earliest branch of the family, (2) the monophyletic Cyprininae emerge next, (3) the barbel-less monophyletic Leuciscinae are a more recent clade with the Phoxinini as the sister group of all other leuciscines, and (4) the Alburninae are polyphyletic and nested within the Leuciscinae. Interrelationships of the Tincinae (*Tinca*), Acheilognathinae (*Rhodeus*), and Gobioninae (*Gobio*) remained unclear as corresponding nodes were generally not statistically supported and results differed from one gene to another. For these reasons, it was necessary to obtain additional sequence data that could lead to improved resolution of cyprinid relationships. This aim was reached through sequencing of the control region of mitochondrial DNA and combination with characters from other genes. Before combination, we checked that the genes did not undergo any "process of discord," in the sense of Maddison (1997), by measuring gene congruence using the incongruence length difference (ILD) test (Farris *et al.*, 1995). When statistically significant character incongruence between two sequence data sets was found, iterative removals of a single taxon followed by ILD tests as described in Lecointre *et al.* (1998) were performed to identify those taxa responsible for incongruence. It was then possible to perform the "partial combination" phylogenetic analysis, from which the

1055-7903/01 $35.00

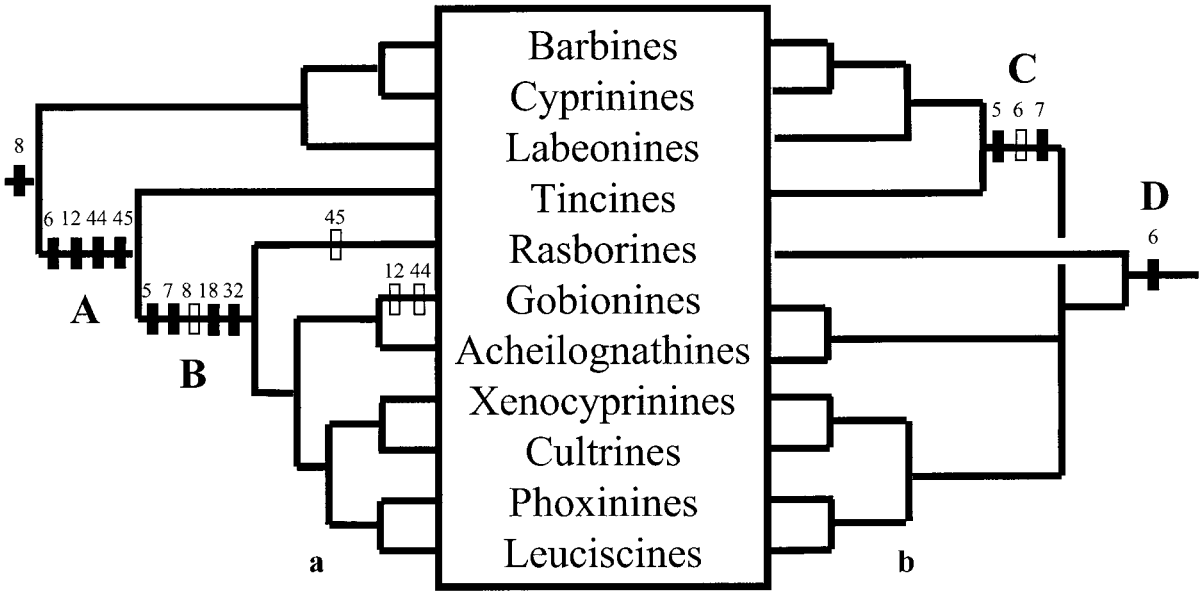


FIG. 1. Two phylogenetic hypotheses for cyprinid intrarelationships based on morpho-anatomical characters. (a) The most parsimonious tree found by Cavender and Coburn (1992) from their own matrix of 47 characters; (b) the most parsimonious tree found by Cavender and Coburn (1992) from the character matrix of Chen *et al.* (1984). A, B, C, D, clades discussed in the text. Character changes are shown for characters for which the interpretation is challenged by our molecular results, i.e., the positions of the rasborines and the tincines. They are numbered according to the matrix of Cavender and Coburn (1992). Empty bars are reversals. These characters (pleisomorphic state; apomorphic state) are 5, crest or blade of the neural complex simple, not divided dorsally; crest of neural complex divided dorsally (forked); 6, fully developed first pleural rib (rib of fifth vertebra) without modified head and parapophysis; rib head and parapophysis of fifth rib modified for greater mobility; 7, pseudobranchial and suprabranchial arteries connected; loss of this connection; 8, interorbital septum formed by orbitosphenoid only; interorbital septum formed by orbitosphenoid and a dorsal component of the parasphenoid; 12, origin of dorsal fin ahead of insertion of pelvic fins; origin of dorsal fin behind pelvic insertion; 18, anterior free supraneural not in contact with neural complex; anterior free supraneural in contact with neural complex; 32, fourth unbranched dorsal rays; three unbranched dorsal rays; 44, posteriorly positioned anal fin suspended from middle section of caudal series; anteriorly positioned anal fin suspended from anterior hemal spines; and 45, anterior maxillary barbel and maxillary foramen present; loss of anterior maxillary barbel and maxillary foramen.

previously identified "incongruent" stretches of DNA sequences were replaced by question marks. This strategy led to two congruent and robust trees: the most parsimonious tree based on the control region sequences and the most parsimonious tree obtained from the partial combination described. These trees confirm previous solid facts but also provide answers to questions previously left open: the rasborines are confirmed as the most basal cyprinid lineage, and *Tinca* is not the sister group of the Cyprininae but the sister group of the clade acheilognathines + gobionines + leuciscines. These relationships are neither those proposed by Chen *et al.* (1984) nor those proposed by Cavender and Coburn (1992).

MATERIALS AND METHODS

Taxonomic Sampling

Sampling was performed as described in Gilles *et al.* (1998). The cyprinids for which we have sequenced the control region comprise 2 rasborines, 6 European cyprinines, the tincine *Tinca tinca*, the European acheilognathine *Rhodeus amarus,* the European gobionine *Gobio gobio,* and 17 European leuciscines. Six cobitoid sequences were used as outgroups. Accession numbers are given in Table 1 and localities are the same as those in Gilles *et al.* (1998).

DNA Extraction, PCR Amplification, and Sequencing

Total DNA was extracted from muscle tissues following the Taberlet and Bouvet (1991) method. A 580-bp section of mtDNA genome from the 16S rRNA gene and a 435-bp section of mtDNA genome from the cytochrome *b* gene were amplified using standard PCR techniques as described in Gilles *et al.* (1998). For the cytochrome *b* gene, two forward primers, depending on the species (details available upon request), NEW-FOR 5′-AGC CTA CGA AAA ACC CAC CC-3′ and LAST-FOR 5′-CTA ATG GCA AGC CTA CGA AN-3′ (Chang *et al.*, 1994), and one universal reverse primer, 34-REV 5′-AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CT CA-3′ (Cantatore *et al.*, 1994), were used. The entire mitochondrial (812- to 1054-bp, depending on the species) control region was also amplified using standard PCR techniques. Amplifications were done using three combinations of primers: (1) ESTFOR 5′-CAT CGG TCT TGT AAT CCG AAG AT-3′ (forward at 120 bases for *Cyprinus carpio*) and NewCRev 5′-GTT TCG GGG TTT GAC AAG GAT A-3′ (reverse at 751 bases), (2) 344F 5′-CTA TTA CTG GCA TCT GGT TCC-3′

TABLE 1

Taxonomic Sampling and GenBank Accession Numbers for Cyprinid Mitochondrial Sequence Data

Taxa	Accession Nos.		
	16S rDNA	D-loop sequences	Cytochrome *b*
Outgroups			
Siluriformes			
Ictaluridae			
Ictalurus melas	AJ247083	—	AJ388468
Characiformes			
Characidae			
Colossoma macropomum	U33616	—	—
Hepsetidae			
Hepsetus odoe	U33992	—	—
Cypriniformes			
Cobitoidea			
Catostomidae			
Catostomus catostomus	—	—	U40554
Cobitidae			
Acanthophtalmus kuhli	AJ247082	AJ388420	AJ388466
Botia sp.	AJ247084	AJ388419	AJ388462
Corbitis taenia	AJ247080	AJ388421	AJ388469
Balitoridae			
Crossotoma lacustre	M91245*	M91245*	M91245*
Neimacheilus barbatulus	—	AJ388422	AJ388467
Gyrinocheilidae			
Gyrinocheilus sp.	AJ247081	AJ388418	AJ388465
Ingroup			
Cyprinoidea			
Cyprinidae			
Rasborinae			
Brachydanio rerio	AF036006	AJ388424	AJ388456
Danio malabaricus	U21384	—	—
Rasbora pauciperforata	L14497	—	—
Rasbora paviei	U21554	—	—
Rasbora trilineata	—	AJ388423	AJ388464
Tanichthys albonubes	U21387	—	—
Cyprininae			
Barbus fluviatilis	AJ247065	AJ388415	AJ388443
Barbus barbus	—	—	AJ388470
Barbus meridionalis 1	AJ247048	AJ388417	AJ388437
Barbus meridionalis 2	AJ247061	AJ388416	AJ388436
Carassius auratus	AJ247070	AJ388413	AJ388458
Cyprinus carpio	X61010*	X61010*	X61010*
Labeo bicolor	AJ247069	AJ388414	AJ388457
Puntius conchonius	L14491	—	AJ388455
Gobioninae			
Gobio gobio 1	AJ247068	AJ388393	—
Gobio gobio 2	AJ247056	AJ388392	AJ388431
Pseudorasbora parva	U21386	—	—
Tincinae			
Tinca tinca	AJ247053	AJ388411	AJ388426
Acheilognathinae			
Rhodeus amarus	AJ247086	AJ388412	AJ388441
Alburninae			
Alburnoides bipunctatus	AJ247072	AJ388408	AJ388427
Alburnus alburnus	AJ247063	AJ388401	AJ388428
Leuciscinae			
Abramis brama	AJ247067	AJ388404	AJ388438
Aspicus aspius	—	—	AJ388460
Blicca bioerkna	AJ247064	AJ388405	AJ388444
Chondrostoma genei	AJ247057	AJ388394	AJ388425
Chondrostoma nasus	AJ247047	AJ388396	AJ388454
Chondrostoma soetta	AJ247060	AJ388397	AJ388452
Chondrostoma toxostoma 1	AJ247040	—	—
Chondrostoma toxostoma 2	AJ247046	AJ388395	AJ388453

TABLE 1—*Continued*

Taxa	Accession Nos.		
	16S rDNA	D-loop sequences	Cytochrome *b*
Chondrostoma toxostoma 3	—	—	AJ388447
Leucaspius delineatus	AJ247071	AJ388402	AJ388459
Leusciscus cabeda	AJ247055	AJ388406	AJ388451
Leuciscus cephalus	AJ247054	AJ388407	AJ388429
Leuciscus leuciscus	AJ247074	—	AJ388449
Leuciscus multicellus 2	AJ247044	AJ388399	AJ388432
Leuciscus soufia 2	AJ247049	AJ388398	AJ388442
Pachychilon pictum	AJ247085	AJ388409	AJ388445
Phoxinus phoxinus	AJ247062	AJ388410	AJ388463
Rutilus rubilio	AJ247059	AJ388400	AJ388440
Rutilus rutilus	AJ247045	—	AJ388439
Scardinius erythrophtalmus	AJ247066	AJ388403	AJ388461
Scardinius erythrophtalmus 2	—	—	AJ388450

Note. —, Not sequenced.
* Complete mitochondrial genome.

(forward at 344 bases) and PHE1R 5′-ACA TCT TCA GTG TTA CGC TT-3′ (reverse at 1026 bases), and (3) CR1F 5′-CCG GGC ATT CTT TTA TAT GC-3′ (forward at 451 bases) and PHE1R 5′-ACA TCT TCA GTG TTA CGC TT-3′ (reverse at 1026 bases). Thermal cycle amplifications were performed in a 50-μl tube as described in Gilles *et al.* (1998). Cycle parameters for the control region primers were as follows: first protocol, 1 min at 92°C (1 cycle); 15 s at 92°C, 45 s at 48°C, 2 min 30 s at 72°C (5 cycles); 15 s at 92°C, 45 s at 52°C, 2 min 30 s at 72°C (30 cycles); 7 min at 72°C (1 cycle); second protocol, 1 min at 92°C (1 cycle); 15 s at 92°C, 45 s at 45°C, 2 min 30 s at 72°C (5 cycles); 15 s at 92°C, 45 s at 48°C, 2 min 30 s at 72°C (30 cycles); 7 min at 72°C (1 cycle).

The amplified DNA segments were purified following the gelase protocol (Epicentre) and then stored at −20°C. Purified fragments were directly sequenced using an automated sequencer (Genome Express S.A.). Sequences were obtained from both forward and reverse primers for each PCR product.

Sequence Analysis

Sequences were entered and aligned using the MUST package (Philippe, 1993). Alignments of the 16S sequence data (58 taxa, 184 positions informative for parsimony) and cytochrome *b* sequence data (45 taxa, 174 positions informative for parsimony) are described elsewhere (Gilles *et al.*, 1998). Saturation was explored for each gene by the plotting of the pairwise number of observed nucleotide differences against the pairwise number of inferred substitutions (Philippe *et al.*, 1994; Hassanin *et al.*, 1998; Lavoué *et al.*, 2000). This was performed by the COMP-MAT program of MUST, the first term being computed by MUST and the second being computed by PAUP 3.1.1. (Swofford, 1993) as the number of steps met in the path joining the two species in the most parsimonious tree. This was obtained by saving the maximum parsimony (MP) tree with its branch lengths from PAUP and transferring it to the AF_PAUP3 program of MUST. Phylogenetic analyses were conducted using heuristic searches of PAUP 3.1.1. Bootstrap proportions (Felsenstein, 1985) were obtained from 1000 iterations using PAUP 3.1.1. The null hypothesis of congruence between gene partitions was tested using the incongruence length difference test (Farris *et al.*, 1995), as performed by the ARNIE program from the Random Cladistics package of Mark Siddall (commands cc-.; mh; bb-; with 1000 iterations; the package is available at http://www.vims.edu/~mes/mes/rchelp.html#arnie).

The ILD test is based on the incongruence length difference as $D_{xy} = L_{(x+y)} - (L_x + L_y)$ (Mickevich and Farris, 1981; Farris *et al.*, 1995), where $L_{(x+y)}$ is the length of the MP tree from the combined analysis and L_x and L_y are the lengths of the MP trees in separate analyses. The following presentation is a simplification of this metric, as in Farris *et al.* (1995). The principle of the ILD test is to test the null hypothesis of congruence between data sets. For this purpose, parsimony analyses are carried out separately for each data set x and y. Then, the lengths of each MP tree obtained are added ($L_x + L_y$) and this length is compared to the sum of lengths ($L_p + L_q$) of the MP trees obtained from two data sets p and q of the same size as the original data sets and generated by random partitioning of the original two data sets. From W random partitions, S is the number of times that $(L_x + L_y) < (L_p + L_q)$ and the null hypothesis of congruence is rejected when the P value of $(1 - S)/(W + 1)$ is small enough, i.e., <5%. The test then indicates that there is more incongruence among the data sets than would be expected from chance alone.

When two given DNA data sets appeared incongruent to each other ($P < 5\%$), visual inspection of the MP trees obtained from each sequence data set rapidly

provided the scope of incongruence. Removal of each species separately followed by a new ILD analysis led to identification of the species responsible for incongruence. Thus, combination of genes excluding these previously identified species can be used for phylogenetic analyses. This approach allowed for combination of congruent data sets in a single phylogenetic analysis (Doyle, 1992, 1997; Maddison, 1997). This strategy fulfills the requirements of both the prior agreement approach (Chippindale and Wiens, 1994; Bull *et al.*, 1993) and the total evidence approach, not in the sense of Kluge (1989) but in the sense of Carnap (1950). Carnap (1950) expressed the "requirement for total evidence" as a principle, not as a procedure (Lecointre and Deleporte, 2000; *Submitted*). He did not specify the method of utilizing all available data but recommended the use of all available knowledge relevant to the hypothesis at hand. In this sense, removing a part of the data that obscures the history of taxa follows the principle of "total evidence" sensu Carnap (1950).

We have reanalyzed the two matrices published by Cavender and Coburn (1992), the first containing 47 anatomical characters polarized by these authors and the second containing 29 anatomical characters interpreted by Chen *et al.* (1984) and reanalyzed by Cavender and Coburn (1992). These analyses were conducted on the taxon sample of the present molecular analyses to allow trees and extra steps comparisons using PAUP 3.1.1 and MacClade (Maddison and Maddison, 1992). From the combined molecular data set, the significance of topological differences between a morphological tree and the molecular MP tree was tested using the Wilcoxon signed-rank test (Templeton, 1983) as performed by PAUP 4 (Swofford, 1998).

RESULTS

Alignments

Alignment of the control region (CR) was achieved within the Cyprinidae, but could not be achieved with confidence along the entire control region between the cyprinid and the outgroup sequences (Cobitoidea). The entire aligned control region of 35 taxa provided 1233 positions among which 614 positions were kept for their nonambiguous alignment between outgroup and ingroup sequences. Portions deleted were 1–39, 106–399 (a region where outgroups are difficult to align and require insertions), 914–922, and 967–1233. This provided 463 variable positions among which 334 were informative for parsimony. This procedure implies a loss of phylogenetic information. To check the phylogenetic information contained in the deleted parts (mainly with regard to the problematic positions of *Gobio, Tinca, Rhodeus,* and leuciscines), a second analysis was conducted. The first study showed that the cyprinine lineage was more basal than other nonrasborine cyprinids. Because the alignment was easier between nonrasborine cyprinids, a longer stretch of the control region was available for investigation of interrelationships of nonrasborine cyprinids by declaration of the subfamily Cyprininae as the outgroup. Thus, a stretch of 1003 positions offered a reliable alignment between 27 nonrasborine cyprinids (deleted portions were 1–29, 106–151, 914–922, and 1081–1233), with 674 variable positions among which 448 were informative for parsimony. This allowed for inference of intraleuciscine relationships with more characters.

Saturation Plots

Absolute saturation analysis of CR transitions and transversions separately showed for both a rather linear plot without any plateau. In Fig. 2 are shown saturation plots for the first set of characters described above including cobitoid outgroups. No mutational saturation was recorded. In the saturation analysis of cytochrome *b* sequences (Appendix 1), transitions at the third codon positions were saturated with superimposed substitutions as shown by the marked plateau, whereas other types of substitutions were not. Transitions at third codon positions of the cytochrome *b* were removed from phylogenetic analysis. In the 16S sequence data, no marked plateau was detected (Appendix 2).

ILD Tests

If the goal of the ILD test is to detect processes of discordance in order to remove from a future combined analysis the gene(s)/taxon(s) that have been subjected to such a process, the best way to proceed is to perform pairwise tests instead of a single global test including the three data sets. Three pairwise ILD tests (with iterative taxon removals) can identify both the taxon and the gene that experienced a process of discordance, whereas a single ILD test on all the data can identify the taxon but not the gene. So, three pairwise tests using the 31 species common to all data sets were performed. The test for incongruence between the cytochrome *b* data and the 16S data yielded a *P* value of 0.19, CR against 16S gave a *P* value of 0.01, and CR against cytochrome *b* gave a *P* value of 0.16. The *P* value obtained from the CR/16S comparison is below the 0.05 threshold, showing that there is more character incongruence between these two data sets than one would expect by chance alone. Thus, one should not combine the complete CR sequence data with the complete 16S sequence data in a simultaneous analysis. Such a statistically significant incongruence suggests that one or several stretches of DNA do not follow the history of the species; a process of discord (Maddison, 1997) must be implied. To identify this stretch of DNA, iterative removals of a single taxon followed by new ILD tests, as in Lecointre *et al.* (1998), detect the sequence responsible for the incongruence. When the species *Pachychilon pictum* was removed, the *P* value increased to above 0.05; testing CR data against 16S

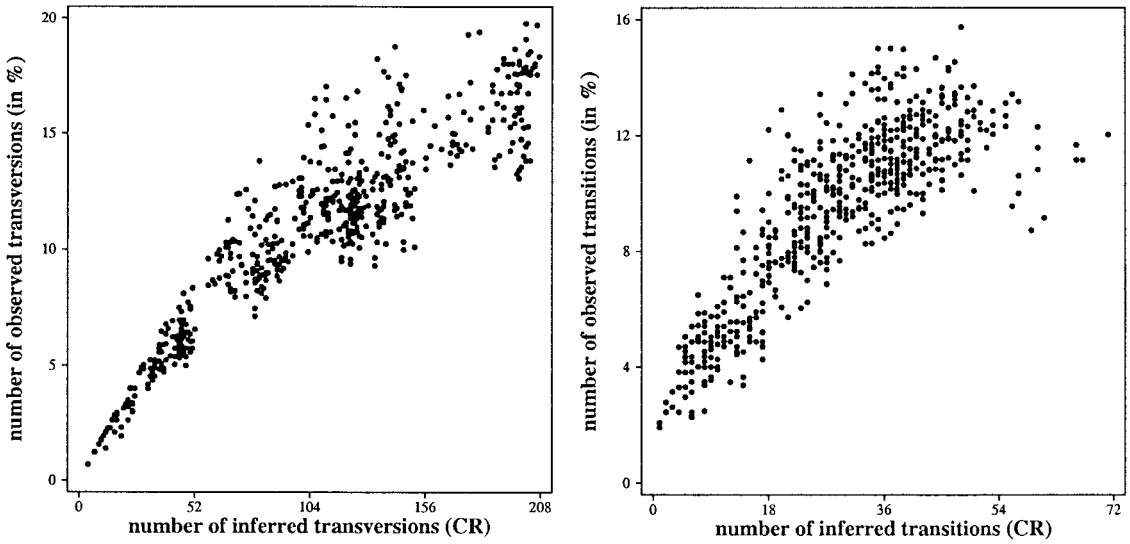


FIG. 2. Saturation plot obtained from the 614 positions of the control region alignable between cyprinid ingroup and cobitoid outgroup taxa. X axis: pairwise number of inferred substitutions; Y axis: pairwise percentage of observed differences.

yielded a *P* value of 0.16. CR and 16S sequences of *P. pictum* were therefore replaced by question marks in the combined analysis of the three genes.

New Data Sets and Trees

There are therefore three data sets. The first is the new control region sequence data alignable with cobitoid outgroups (35 taxa, 614 characters, MP tree of Fig. 3). The second is a longer stretch of the new control region sequence data with the Cyprininae taken as outgroup (27 taxa, 1003 characters, MP tree of Fig. 4). The third is the combined analysis of CR, 16S, and cytochrome *b* genes (31 taxa, 1374 characters, MP tree of Fig. 5) generated by removal of the 16S and CR sequences of *P. pictum* (because of incongruence) and removal of transitions at the third codon positions of the cytochrome *b* (because of mutational saturation). These trees are markedly congruent and far better resolved and robust than previously published cyprinid molecular trees (Gilles *et al.*, 1998; Briolay *et al.*, 1998). One should note that the trees based on the control region (Figs. 3 and 4) and the tree based on the partial combination (Fig. 5) are not congruent with respect to the relative systematic positions of *Rhodeus* and *Gobio*. However, this conflict is not strong because none of these taxa are responsible for statistically significant character incongruence, and the nodes where the contradictions appear are not robust; one could consider these nodes collapsed.

The main phylogenetic conclusions of these three molecular analyses are as follows. The rasborine subfamily is the most basal subfamily within the Cyprinidae (Figs. 3 and 5). The monophyletic Cyprininae emerges next. In the three analyses, *T. tinca* is the sister group of all the remaining nonrasborine and noncyprinine species (Figs. 3–5). This result is here more robust than in previous studies in which the position of *Tinca* was ambiguous (Briolay *et al.*, 1998; Gilles *et al.*, 1998; Zardoya and Doadrio, 1999). In the two trees based on the control region, there is a contradiction in the relative positions of *Rhodeus* and *Gobio*. However, none of the corresponding nodes is supported. In the tree based on the partial combination (Fig. 5), *Rhodeus* emerges first and then *Gobio*. *Gobio* is therefore the sister group of the Leuciscinae, in which the Phoxinini is the sister group of the Leuciscini, as suggested in Briolay *et al.* (1998). This does not strongly contradict the relationships of *Phoxinus* inferred in Gilles *et al.* (1998), in which *Phoxinus* is the sister group of *Rhodeus* but with very low support. Within the Leuciscinae, the genus *Leuciscus* is paraphyletic, with *L. multicellus* and *L. souffia* close to *Chondrostoma* and *L. cephalus* and *L. cabeda* closer to the clade *Abramis* + *Blicca* + *Leucaspius* + *Alburnus*. Moreover, we confirm that the Alburninae are paraphyletic.

Comparing Trees

When our molecular topology (Fig. 5) is fitted to the Cavender and Coburn (1992) matrix, the tree is 71 steps long, 12.7% longer than the original morphological MP tree (63 steps including autapomorphies). When our molecular topology is fitted to the Chen *et al.* (1984) matrix, the tree is 39 steps long, 11.4% longer than the original morphological tree length (35 steps). Our molecular data therefore seem to contradict the two morphological matrices to the same extent. The Cavender and Coburn (1992) topology (Fig. 1a) was applied to the molecular combined data set. The tree length increased from 2591 steps (which is the length of the MP tree of Fig. 5) to 2612. Using a 5% threshold, the Wilcoxon signed-rank test (Templeton, 1983) per-

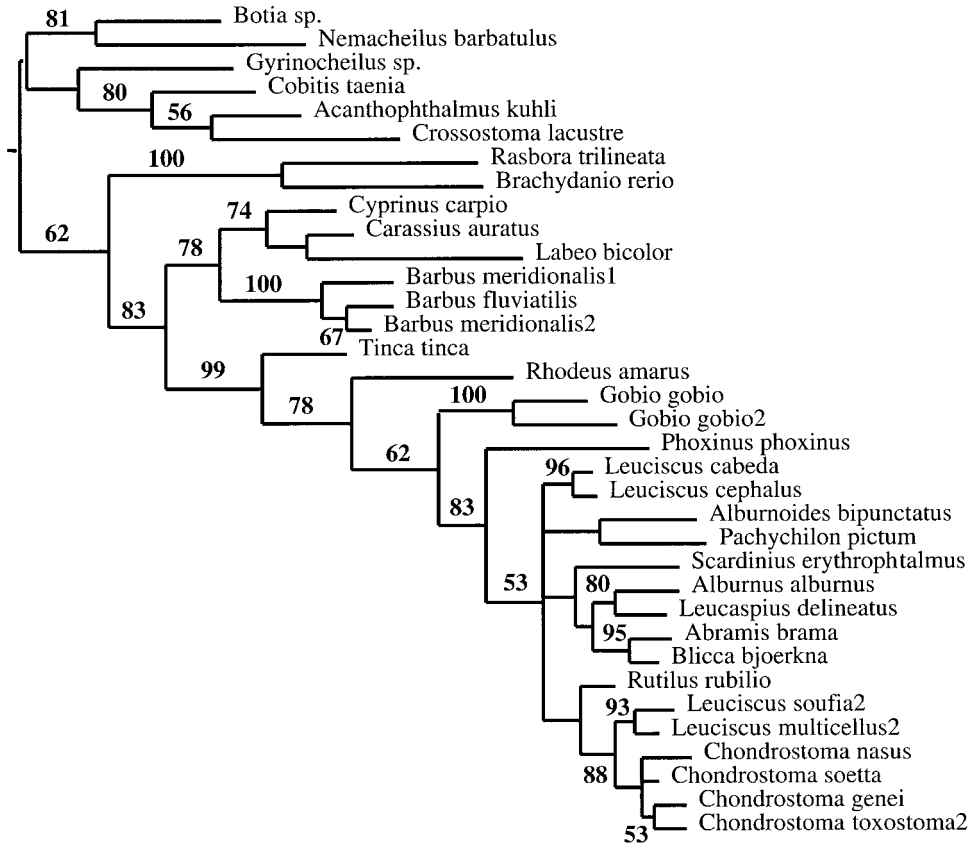


FIG. 3. Strict consensus of the four equiparsimonious trees found from a parsimony analysis (general heuristic search of PAUP 3.1.1.) of the matrix of 614 positions in the control region alignable between cyprinid ingroup and cobitoid outgroup taxa. Branch lengths (ACCTRAN) are shown on the consensus. Each equiparsimonious tree is 1899 steps long with a consistency index of 0.470 and a retention index of 0.571. Numbers on branches are bootstrap proportions obtained for the node using 1000 replicates.

formed using PAUP 4 (Swofford, 1998) led to the rejection of the null hypothesis that the molecular tree (Fig. 5) is the same as the tree constrained according to the topology of Fig. 1a ($P = 0.0103$). The Chen *et al.* (1984) topology (Fig. 1b) was also applied to the molecular combined data set. The tree length increased from 2591 steps (Fig. 5) to 2638. Using a 5% threshold, the Wilcoxon signed-rank test also led to the rejection of the null hypothesis that the molecular tree (Fig. 5) is the same as the tree constrained according to the topology of Fig. 1b ($P < 0.0001$). The most parsimonious molecular tree obtained from the combination designed as explained above (Fig. 5) is therefore significantly different from each of the two cladograms based on morphological characters. A close look at the morphological characters is thus justified.

DISCUSSION

It is striking to notice that the tree of Cavender and Coburn (1992) differs from the tree of Chen *et al.* (1984; as retrieved by Cavender and Coburn, 1992) by the systematic positions of the rasborines and the tincines (Fig. 1), exactly the same points of main disagreement with the molecular trees. Our molecular trees agree with Cavender and Coburn (1992) with regard to the position of the tincines, but not to the position of the rasborines, and agree with Chen *et al.* (1984) with regard to the position of the rasborines, but not to the position of the tincines.

With such localized incongruence between molecular and morphological data, the reliability of both types of characters must be evaluated. The reliability of molecular characters has already been tested by saturation and ILD tests. The quality of those anatomical synapomorphies that are contradicted by molecular data is now under scrutiny. In the following, anatomical characters that provide synapomorphies for the clade acheilognathines + gobionines will not be discussed because this clade is not strongly challenged by molecular data: the conflicting node in our molecular trees is never supported with high bootstrap proportions. We will discuss the changes in the interpretation of the evolution of morphological characters imposed by the molecular positions of the rasborines and tincines. The molecular position of the rasborines contradicts four synapomorphies of the Cavender and Coburn (1992) tree at node A (Fig. 1a; characters 6, 12, 44, and 45)

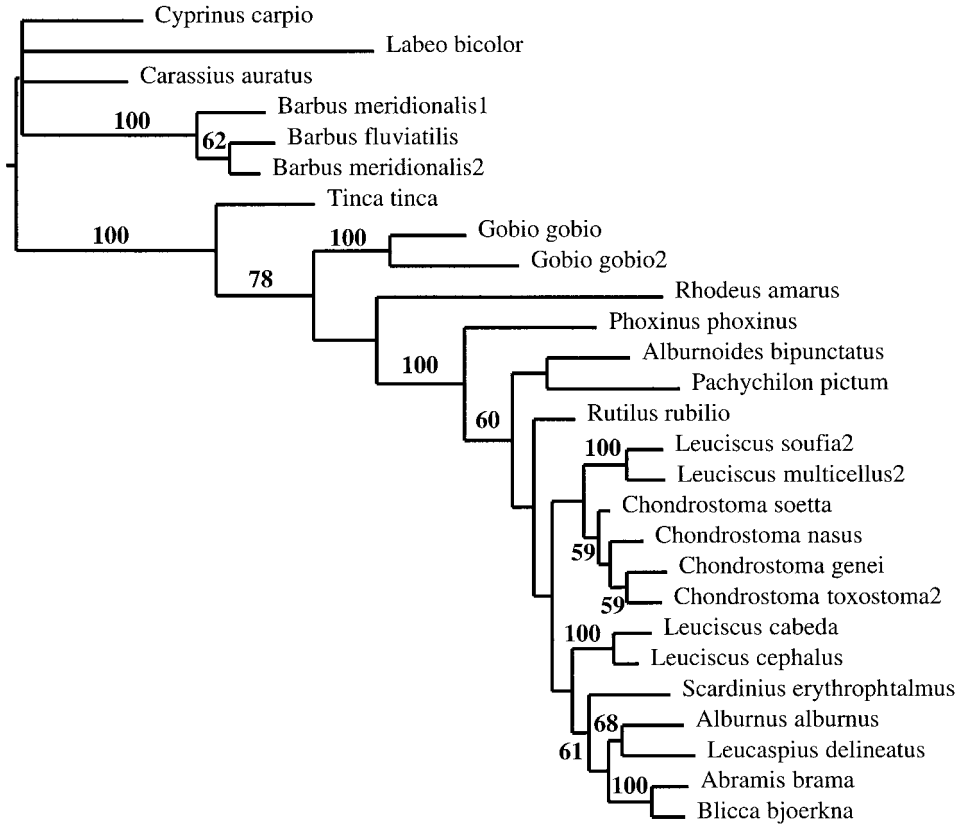


FIG. 4. Strict consensus of the three equiparsimonious trees found from a parsimony analysis (general heuristic search of PAUP 3.1.1.) of the matrix of 1003 positions in the control region alignable between nonrasborine cyprinids. Following the tree in Fig. 3, the Cyprininae are taken as the outgroup. Branch lengths (ACCTRAN) are shown on the consensus. Each equiparsimonious tree is 2017 steps long with a consistency index of 0.566 and a retention index of 0.603. Numbers on branches are bootstrap proportions obtained for the node using 1000 replicates.

and five synapomorphies at node B (Fig. 1a; characters 5, 7, 8, 18, and 32). Only two of these characters seem to be reliable and difficult to contradict, whereas the others are rather likely to change and difficult to polarize.

Characters 12 and 44 concern the positions of the dorsal and anal fins, respectively. These are characters linked to swimming habits that are driven by the ecological niches occupied by fishes, possibly of diverse origins. The relative positions of these fins are rather susceptible to convergence at the large taxonomic scope needed for outgroup comparison. Character 45 does not contradict the molecular tree. Presence of two pairs of maxillary barbels is considered the primitive condition for cyprinids; the loss of the anterior maxillary barbels is a derived state. This loss occurred at node A (Fig. 1), implying a reversal for rasborines which possess two pairs of barbels. A basal position of rasborines would imply no extra step for this character. The synapomorphy at node A given by character 6 is difficult to reject. The primitive state for cyprinids is the unmodified first pleural rib and its unmodified parapophysis of the fifth vertebra. This condition is found in ostariophysan outgroups. The derived state found in clade A (Fig. 1) consists of modified size and morphology of both the articulating head of the first pleural rib and the parapophysis on which the rib articulates, resulting in an angle different from that of the pleural ribs that follow (Cavender and Coburn, 1992, p. 303). According to our molecular trees, this condition would have arisen twice in cyprinid evolution, once at node A and once in the basal rasborine lineage.

Node B (Fig. 1) is supported by two linked synapomorphies (characters 5 and 18), the forked crest of the neural complex and the arrangement of the free supraneural and the neural complex. According to Cavender and Coburn (1992), the primitive condition of character 5 for the cyprinids is the blade-like crest of the neural complex, which is the opposite of the interpretation of Chen *et al.* (1984). This character is therefore of limited reliability. The same doubts can be expressed for character 32 (the primitive state is four or more unbranched dorsal fin rays; the derived state is three), as the polarization proposed by Cavender and Coburn (1992) is opposite of that presented by Gosline (1978). In addition, this character frequently changes among the Leuciscinae, reducing its utility. Character 7, the absence of connection between the pseudobranchial

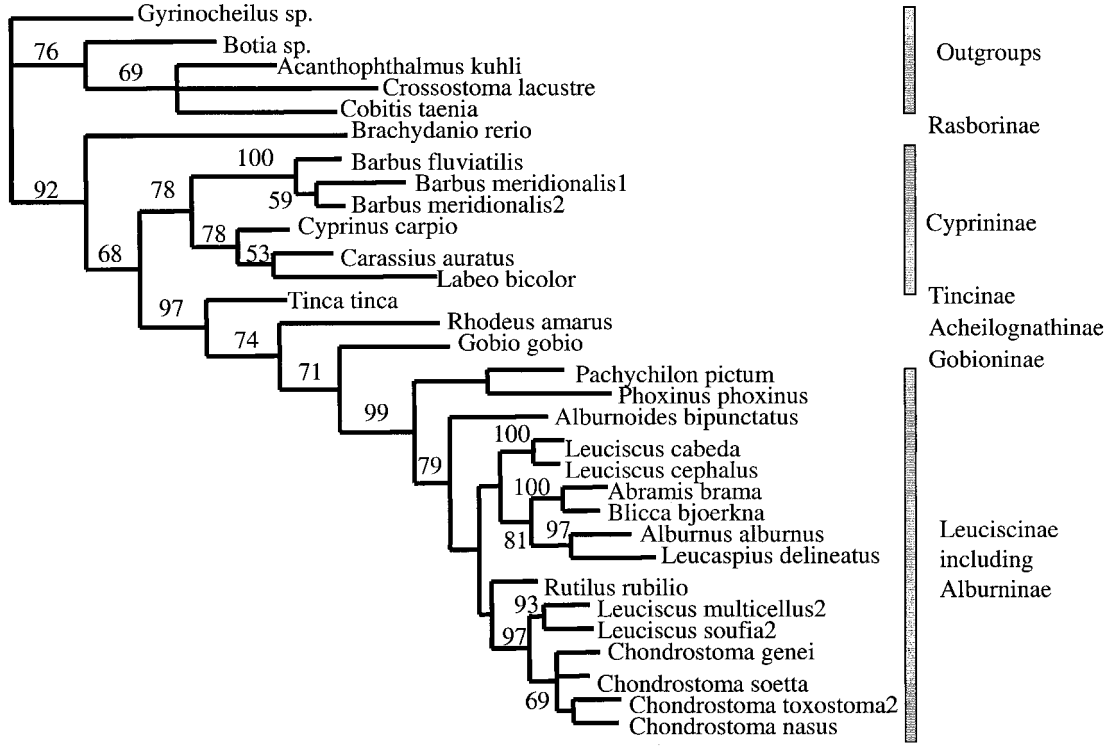


FIG. 5. Strict consensus of the six equiparsimonious trees found from a parsimony analysis (heuristic search of PAUP 3.1.1.) of the partial combination of three genes (16S rDNA, control region, and cytochrome *b*), providing 1374 positions alignable between cyprinid ingroup and cobitoid outgroup taxa. Branch lengths (ACCTRAN) are shown on the consensus. Each equiparsimonious tree is 2591 steps long with a consistency index of 0.480 and a retention index of 0.565. Numbers on branches are bootstrap proportions obtained for the node using 1000 replicates. Transitions at the third codon position have been removed from the analysis because of mutational saturation (see Appendix 1), and the 16S and control region sequences of the species *Pachychilon pictum* have been replaced by question marks in the matrix because of statistically significant incongruence provoked by these sequences.

and the suprabranchial arteries, has a low confidence according to Cavender and Coburn (1992). Only character 8, the interorbital septum formed by the orbitosphenoid bone only, seems reliable. Our molecular trees would imply that this septum formed by the orbitosphenoid and a dorsal component of the parasphenoid evolved twice in cyprinids.

The position of the Tincines in the molecular topology contradicts three synapomorphies in the Chen *et al.* (1984) tree at node C (Fig. 1b; characters 5–7). The lack of reliability of characters 5 and 7 has already been discussed above. Character 6 has a consistency index of 0.5 in the Chen *et al.* (1984) tree: the first pleural rib is unmodified in outgroups, modified in rasborines (a change is needed in the basal cyprinid node D; Fig. 1b), and unmodified in clade C (implying a reversal in node C). Therefore, a change of position of Tincinae inferred from the molecular data would not change the length of the tree. However, Cavender and Coburn (1992) coded a question mark for the state of this character found in *T. tinca*.

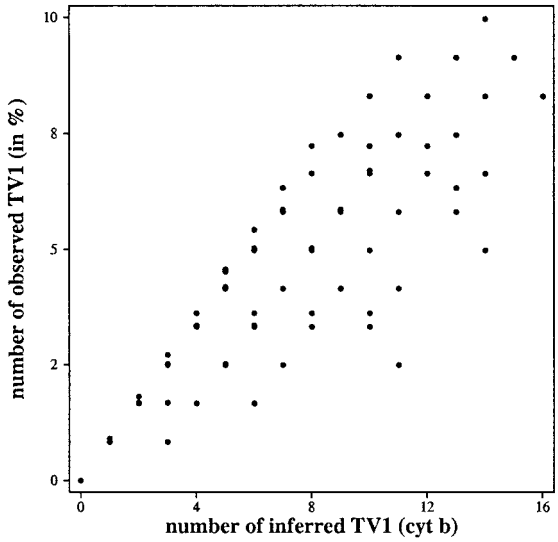
Three reliable synapomorphies are therefore challenged by the present molecular trees, two in the tree of Fig. 1a and one in the tree of Fig. 1b. They represent two characters that may need to be reinterpreted: the modification of the first pleural rib and its parapophysis and the composition of the interorbital septum. Otherwise, we must admit that they could have arisen twice in cyprinid evolution.

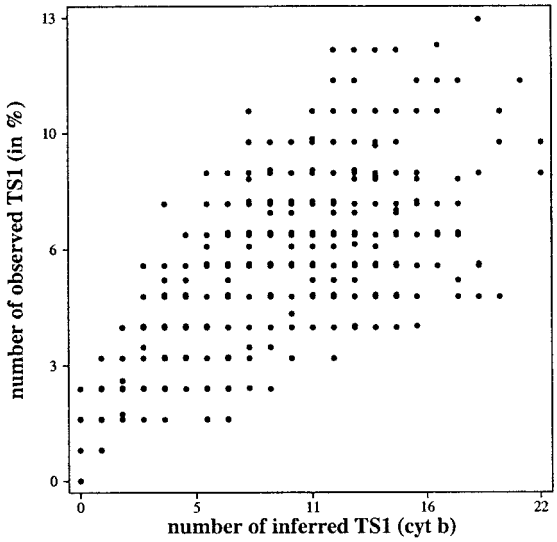
To conclude, more accurate cyprinid intrarelationships are now available. The rasborine subfamily is the most basal subfamily within the Cyprinidae. The monophyletic Cyprininae emerges next. *T. tinca* is the first sister group of all the remaining nonrasborine and noncyprinine species; then *Rhodeus* emerges. *Gobio* is the sister group of the Leuciscinae, in which the Phoxinini is the sister group of the Leuciscini. The genus *Leuciscus* (within the Leuciscinae) and the subfamily Alburninae are both clearly paraphyletic. Two anatomical characters need to be reinterpreted or considered as homoplastic in cyprinid evolution: the modification of the first pleural rib and its parapophysis and the bone composition of the interorbital septum.

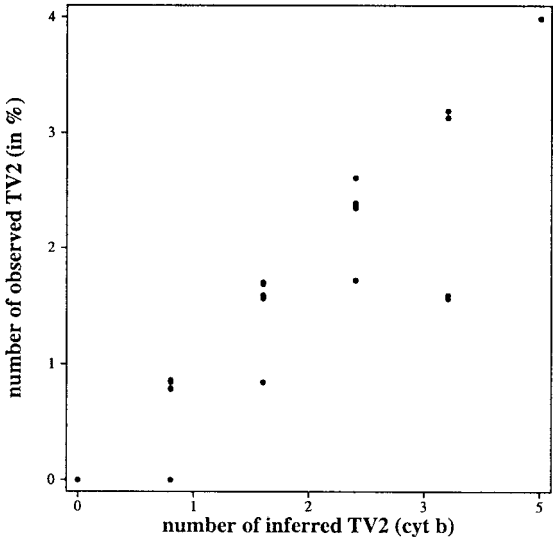
APPENDIX 1

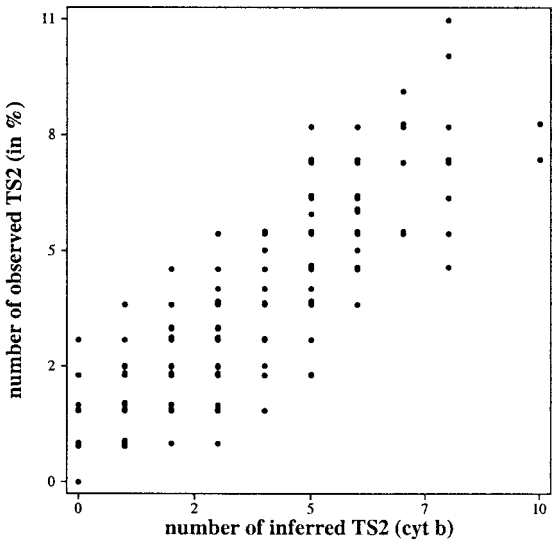
Saturation plots obtained from 369 positions of the cytochrome *b* of 45 cyprinids and cobitoid outgroups. X axis: pairwise number of inferred substitutions; Y axis: pairwise percentage of observed differences. Left

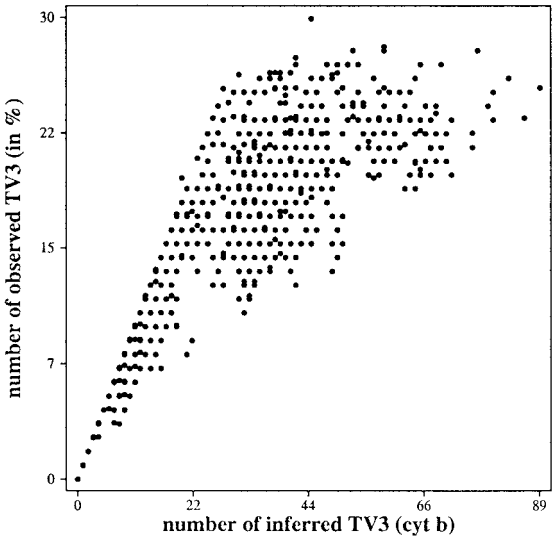
APPENDIX 1

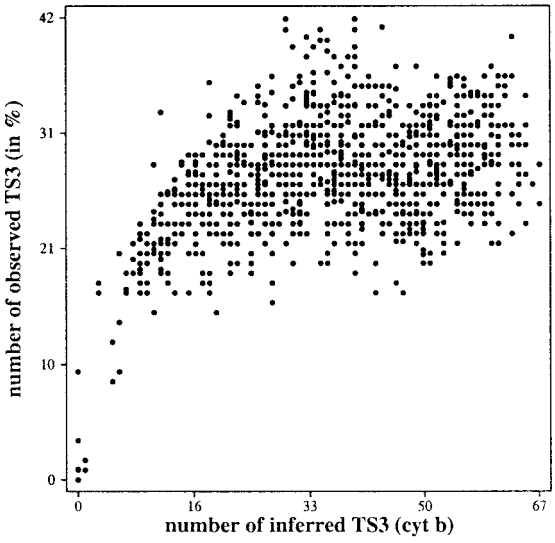

number of observed TV1 (in %)
number of inferred TV1 (cyt b)
0
2
5
8
10
0
4
8
12
16


number of observed TS1 (in %)
number of inferred TS1 (cyt b)
0
3
6
10
13
0
5
11
16
22

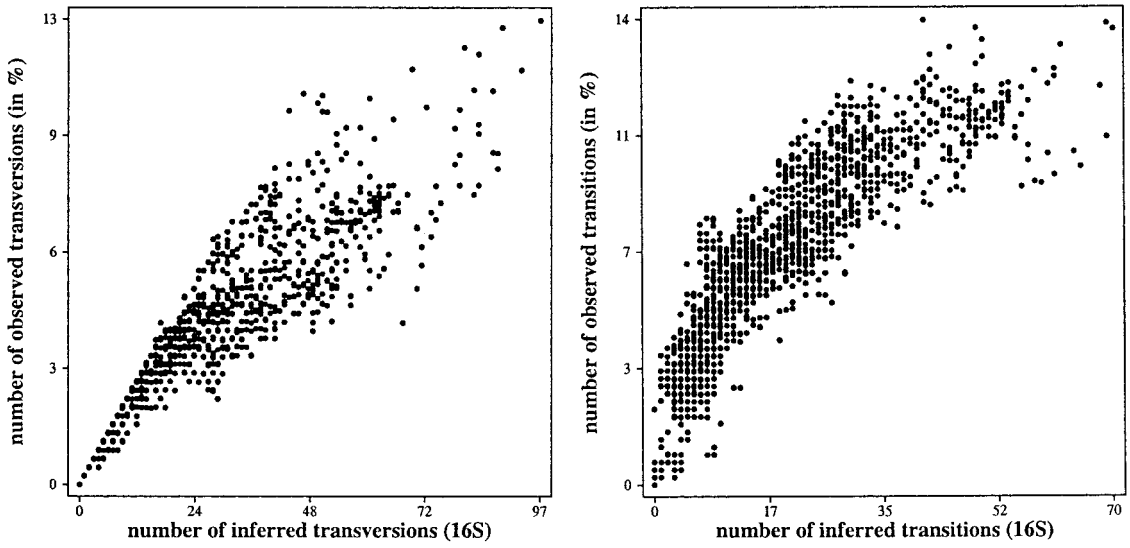

number of observed TV2 (in %)
number of inferred TV2 (cyt b)
0
1
2
3
4
0
1
2
3
5


number of observed TS2 (in %)
number of inferred TS2 (cyt b)
0
2
5
8
11
0
2
5
7
10


number of observed TV3 (in %)
number of inferred TV3 (cyt b)
0
7
15
22
30
0
22
44
66
89


number of observed TS3 (in %)
number of inferred TS3 (cyt b)
0
10
21
31
42
0
16
33
50
67

APPENDIX 2



(transversions; TV) to right (transitions; TS): TV1, TS1, TV2, TS2, TV3, TS3. Typographic errors of Gilles *et al.* (1998) have been corrected.

APPENDIX 2

Saturation plots obtained from 457 positions of the 16S mitochondrial DNA of 58 cyprinids and cobitoid outgroups. X axis: pairwise number of inferred substitutions; Y axis: pairwise percentage of observed differences. Left: transversions; right: transitions. Typographic errors of Gilles *et al.* (1998) have been corrected.

REFERENCES

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

Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L., and Wadell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* **42:** 384–397.

Cantatore, P., Roberti, M., Pesole, G., Ludovico, A., Milella, F., Gadaleta, M., and Saccone, C. (1994). Evolutionary analysis of cytochrome b sequences in some Perciformes: evidence for a slower rate of evolution than in mammals. *J. Mol. Evol.* **39:** 589–597.

Carnap, R. (1950). "Logical Foundations of Probability," Univ. of Chicago Press, Chicago.

Cavender, T. M., and Coburn, M. M. (1992). Phylogenetic relationships of North American Cyprinidae. *In* "Systematics, Historical Ecology and North American Freshwater Fishes" (R. L. Mayden, Ed.), pp. 293–327. Stanford Univ. Press, Stanford, CA.

Chang, Y., Huang, F., and Lo, T. (1994). The complete nucleotide sequence and gene organisation of carp (*Cyprinus carpio*) mitochondrial genome. *J. Mol. Evol.* **38:** 138–155.

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

Chippindale, P. T., and Wiens, J. J. (1994). Weighting, partitioning, and combining characters in phylogenetic analysis. *Syst. Biol.* **43:** 278–287.

Doyle, J. J. (1992). Gene trees and species trees: Molecular systematics as one-character taxonomy. *Syst. Bot.* **17:** 144–163.

Doyle, J. J. (1997). Trees within trees: Genes and species, molecules and morphology. *Syst. Biol.* **46:** 537–553.

Farris, J. S., Källersjö, M., Kluge, A. G., and Bult, C. (1995). Testing significance of incongruence. *Cladistics* **10:** 315–319.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39:** 783–791.

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

Gosline, W. A. (1978). Unbranched dorsal fin rays and subfamily classification of the fish family Cyprinidae. *Occas. Pap. Mus. Zool. Univ. Michigan* **684:** 1–21.

Hassanin, A., Lecointre, G., and Tillier, S. (1998). The "Evolutionary Signal" of homoplasy in protein coding gene sequences and its consequences for a priori weighting in phylogeny. *C. R. Acad. Sci.* **321:** 611–620.

Howes, G. (1991). Systematics and biogeography: An overview. *In* "Cyprinid Fishes: Systematic Biology and Exploitation" (I. Winfield and J. Nelson, Eds.), pp. 1–54, Chapman & Hall, New York.

Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* **38:** 7–25.

Lavoué, S., Bigorne, R., Lecointre, G., and Agnèse, J. F. (2000). Phylogenetic relationships of mormyrid electric fishes (Mormyridae; Teleostei) inferred from cytochrome *b* sequences. *Mol. Phylogenet. Evol.* **14:** 1–10.

Lecointre, G., and Deleporte, P. (2000). Le Principe de "Total Evidence" requiert l'exclusion de données trompeuses. *In* "Caractères" (V. Barriel and T. Bourgoin, Eds.), pp. 129–151. Biosystema 18, Publication de la Société Française de Systématique, Paris, France.

Lecointre, G., Rachdi, L., Darlu, P., and Denamur, E. (1998). *Escherichia coli* molecular phylogeny using the Incongruence Length Difference test. *Mol. Biol. Evol.* **15:** 1685–1695.

Maddison, W. P. (1997). Gene trees in species trees. *Syst. Biol.* **46:** 523–536.

Maddison, W. P., and Maddison, D. R. (1992). "MacClade: Analysis of Phylogeny and Character Evolution." Version 3.01. Sinauer, Sunderland, MA.

Mickevich, M. F., and Farris, J. S. (1981). The implications of congruence in *Menidia*. *Syst. Zool.* **30:** 351–370.

Philippe, H. (1993). MUST: A computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res.* **21:** 5264–5272.

Philippe, H., Sorhannus, U., Baroin, A., Perasso, R., Gasse, F., and Adoutte, A. (1994). Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. *J. Evol. Biol.* **7:** 247–265.

Siddall, M. E. Random Cladistics, User's manual, University of Toronto, Zoology Department (http://www.vims.edu/~mes/mes/rchelp.html#arnie).

Swofford, D. L. (1993). Phylogenetic analysis using parsimony (PAUP) 3.1.1. Illinois Natural History Survey, Champaign, IL.

Swofford, D. L. (1998). Phylogenetic analysis using parsimony (PAUP) Version 4.0, Sinauer, Sunderland, MA.

Taberlet, P., and Bouvet, J. (1991). A single plucked feather as a source of DNA for bird genetic studies. *Auk* **108:** 58.

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.

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