Phylogenetic studies of sinipercid fish (Perciformes: Siniperidae) based on multiple genes, with first application of an immune-related gene, the virus-induced protein (viperin) gene

Dali Chen\textsuperscript{a,b,1}, Xianguang Guo\textsuperscript{c,1}, Pin Nie\textsuperscript{b}

\textsuperscript{a}Department of Parasitology, School of Preclinical and Forensic Medicine, Sichuan University, Chengdu, Sichuan Province 610041, China
\textsuperscript{b}State Key Laboratory of Freshwater Ecology and Biotechnology, Laboratory of Fish Diseases, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei Province 430072, China
\textsuperscript{c}Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan Province 610041, China

\textbf{A R T I C L E  I N F O}

Article history:
Received 13 November 2009
Revised 25 January 2010
Accepted 30 January 2010
Available online 4 February 2010

Keywords:
Phylogeny
Siniperidae
Siniperca
Coreoperca
Viperin gene
57 ribosomal protein gene

\textbf{A B S T R A C T}

The sinipercid fish represent a group of 12 species of freshwater percoid fish endemic to East Asia. To date published morphological and molecular phylogenetics hypotheses of sinipercid fish are part congruent, and there are some areas of significant disagreement with respect to species relationships. The present study used separate and combined methods to analyze 7307 bp of data from three mitochondrial genes (cyt b, CO1 and 16S rRNA; \textasciitilde 2312 bp) and three nuclear genes (viperin, the first two introns of 57 ribosomal protein gene; \textasciitilde 4995 bp) for the attempts to estimate the relationships among sinipercids and to assess the phylogenetic utility of these markers. Phylogenetic trees were reconstructed using maximum parsimony, maximum likelihood and partitioned Bayesian analyses. Despite the detection of significant heterogeneity of phylogenetic signal between the mitochondrial and nuclear partitions, the combined data analysis represented the best-supported topology of all data. The sinipercid fish form a monophyletic group with two distinct clades, one corresponding to the genus Siniperca and the other to Coreoperca. Coreoperca whiteheadi is the sister taxon to Coreoperca herzi plus Coreoperca kawamebari. In the Siniperca, Siniperca undulata is the sister taxon to the other members of Siniperca, within the subclade containing the other members of the genus, Siniperca chuatsi and Siniperca kneri are sister species, next joined by Siniperca obscura, Siniperca roulei, Siniperca scherzeri and finally by Siniperca fortis. The potential utilities of six different genes for phylogenetic resolution of closely related sinipercid species were also evaluated, with special interest in that of the novel virus-induced protein (viperin) gene.

\textcopyright 2010 Elsevier Inc. All rights reserved.

1. Introduction

The sinipercids, a group of freshwater percoid fish, are endemic in China, Vietnam, Japan and Korea, with the majority of species recorded in China (Zhou et al., 1988). A total of twelve species have been described in this group, and disputes over their taxonomy and phylogeny have ever been solved yet. These fish were once assigned to only one genus, Siniperca (e.g., Zheng, 1989), or to two genera Siniperca, Coreoperca (e.g., Liu and Chen, 1994; Nelson, 2006), or even three genera, Siniperca, Coreoperca, Coreosiniperca (e.g., Kong and Zhou, 1993). Moreover, the family to which these fish belong is also a dispute (e.g., Johnson, 1984; Waldman, 1986), although Nelson (2006) considered that these fish can be allocated into Centropomidae, Percichthyidae, or, as is more appropriate and done by Roberts (1993), in their own family Siniperidae.

Despite the taxonomical disputes, it is a bit surprising that relatively few studies have been carried out in relation to the phylogeny of sinipercids. Using morphological characters and allozymes, Kong and Zhou (1992) considered that sinipercids are of monophyletic origin, confirming the identification of two genera, Siniperca and Coreoperca. On the basis of morphology, other authors (e.g., Liu and Chen, 1994; Roberts, 1993; Yabumoto and Uyeno, 2000) also suggested that sinipercids should be clarified into the two genera. Much recently, Shirai et al. (2003), using mitochondrial cyt b gene sequences, analyzed the phylogenetic relationship of five species of sinipercids, and concluded that these fish were monophyletic with the validity of the two genera. Recently, Zhao et al. (2005, 2006a,b) showed the same conclusion, by using sequences of 16S rRNA, cyt b and of mitochondrial control region, but they provided some additional evidence on Coreosiniperca, which should be merged into Siniperca. It is rather obvious that all these authors, while studying the phylogeny considered that...
sinipercid fish should be clarified into an independent family, Siniper-cidae. However, also using cyt b but incorporated with many other cyt b sequences from a wide range of perciform fish, Chen et al. (2007) found the non-monophyly of these sinipercids, with obviously no comments on their family status. In addition, Chen et al. (2007) also demonstrated that this group does not seem to have a very clear phylogenetic history, for different methods of phylogenetic reconstruction result in different tree topologies.

No nuclear DNA sequences, however, have been employed to test the phylogenetic relationship of sinipercid fish. Single- or low-copy nuclear genes which may represent a source of multiple, unlinked and independently evolving loci, may provide ideal data-phylogenetic reconstruction result in different tree topologies.

The total genomic DNA was extracted from ethanol-preserved muscle following the method of Sambrook et al. (1989). The CO1 gene (≈666 bp) was amplified with the primers FishF1 and FishR1 as described by Ward et al. (2005). One cyt b gene sequence in Coreoperca herzi was amplified with the primers L14724 and H15915 as reported by Xiao et al. (2001), and other cyt b genes were reported in a previous paper (Chen et al., 2007). To sequence the first two introns of the S7, the primers used by Chow and Hazama (1998) were employed in this study. The 16S rRNA gene (≈580 bp) was obtained with the universal primers 16Sar-L and 16Sbr-H (Palumbi, 1996). The PCR cycling conditions were 94°C for 3 min followed by 30 cycles of 94°C for 30 s, 55–63°C (adjusted according to the quality of template DNA) for 30 s and 72°C for 1 min, and then a final elongation step at 72°C for 10 min. The amplified DNA fragments were purified via spin columns and sequenced with an ABI 3730 automated DNA sequencer following the manufacturer’s protocol.

Viperin gene sequences were constructed by linking together overlapping segments amplified with three primers designed on the basis of Sun and Nie (2004). The primer positions and the positions in the complete viperin gene of Siniperca chuatsi are listed in Table 2 and Fig. 1, respectively. The PCR cycling conditions for primers 1 (G1f and G1r) and 2 (G2f and G2r) were 94°C for 5 min followed by 35 cycles of 94°C for 30 s, 63°C for 1 min and 72°C for 1 min 30 s, and then a final elongation step at 72°C for 10 min. The PCR cycling conditions for primers 3 (G3f and G3r) were 94°C for 5 min followed by 35 cycles of 94°C for 30 s, 54°C for 1 min and 72°C for 1 min, and then a final elongation step at 72°C for 10 min. Each viperin segment amplified as a single fragment in all taxa, and an initial screening for the presence of heterozygous sites was done by cloning PCR products into PMD18-T vector (TAKARA). The first segment of viperin (amplified with G1f and G1r, ≈1039 bp) was sequenced with the universal primers M13 for five clones each of 4 genera (Siniperca, Coreoperca, Niphon and Lateolabrax). No evidence of heterozygosity was found, and all subsequent sequencing was performed with one clone of a segment in each taxon. Sequences were then determined in both directions for each species and submitted for BLAST searching (Altschul et al., 1997) in GenBank to verify the data, with the accession numbers listed in Table 1.

### Table 1

<table>
<thead>
<tr>
<th>Family</th>
<th>Species</th>
<th>Origin</th>
<th>Viperin</th>
<th>S7 intron 1</th>
<th>S7 intron 2</th>
<th>Cyt b</th>
<th>CO1</th>
<th>16S rRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siniperidae</td>
<td>Siniperca chuatsi</td>
<td>Changde, China</td>
<td>AY395718a</td>
<td>DQ864745</td>
<td>EF143358</td>
<td>EF143402</td>
<td>EF143387</td>
<td>AY898947b</td>
</tr>
<tr>
<td></td>
<td>S. kneri</td>
<td>Changde, China</td>
<td>EF143401</td>
<td>DQ864749</td>
<td>EF143365</td>
<td>EF143407</td>
<td>EF143389</td>
<td>AY898948b</td>
</tr>
<tr>
<td></td>
<td>S. fortis</td>
<td>Shihiting, China</td>
<td>EF143400</td>
<td>DQ864747</td>
<td>EF143361</td>
<td>EF143405</td>
<td>EF143388</td>
<td>EF143378</td>
</tr>
<tr>
<td></td>
<td>S. obscura</td>
<td>Cili, China</td>
<td>EF143402</td>
<td>DQ864750</td>
<td>EF143364</td>
<td>EF143408</td>
<td>EF143390</td>
<td>DQ435335b</td>
</tr>
<tr>
<td></td>
<td>S. scherzeri 1</td>
<td>Changde, China</td>
<td>EF143403</td>
<td>DQ864751</td>
<td>EF143366</td>
<td>EF143403</td>
<td>EF143391</td>
<td>AY898949b</td>
</tr>
<tr>
<td></td>
<td>S. scherzeri 2</td>
<td>Changju, South Korea</td>
<td>EF143404</td>
<td>DQ864752</td>
<td>EF143367</td>
<td>EF143404</td>
<td>EF143392</td>
<td>EF143380</td>
</tr>
<tr>
<td></td>
<td>S. undulata</td>
<td>Lianzhou, China</td>
<td>EF143405</td>
<td>DQ864753</td>
<td>EF143363</td>
<td>EF143406</td>
<td>EF143393</td>
<td>DQ435334b</td>
</tr>
<tr>
<td></td>
<td>S. roulei</td>
<td>Changde, China</td>
<td>EF143394</td>
<td>DQ864736</td>
<td>EF143360</td>
<td>EF143394</td>
<td>EF143381</td>
<td>DQ435333b</td>
</tr>
<tr>
<td>Coreoperca herzi</td>
<td>South Korea</td>
<td>EF143395</td>
<td>DQ864737</td>
<td>EF143370</td>
<td>EF864731</td>
<td>EF143382</td>
<td>EF143374</td>
<td>EF143374</td>
</tr>
<tr>
<td>C. kawamebari</td>
<td>Niigata, Japan</td>
<td>EF143396</td>
<td>DQ864738</td>
<td>EF143371</td>
<td>EF143401</td>
<td>EF143383</td>
<td>EF143375</td>
<td>EF143375</td>
</tr>
<tr>
<td>C. whiteheadi</td>
<td>Changde, China</td>
<td>EF143397</td>
<td>DQ864740</td>
<td>EF143368</td>
<td>EF143405</td>
<td>EF143384</td>
<td>AY898950b</td>
<td></td>
</tr>
<tr>
<td>Niphon spinosus</td>
<td>Fukuoka, Japan</td>
<td>EF143399</td>
<td>DQ864743</td>
<td>EF143355</td>
<td>EF143403</td>
<td>EF143386</td>
<td>EF143377</td>
<td>EF143377</td>
</tr>
<tr>
<td>Lateolabrax maculatus</td>
<td>Guangzhou, China</td>
<td>EF143398</td>
<td>DQ864742</td>
<td>EF143357</td>
<td>EF143404</td>
<td>EF143385</td>
<td>EF143376</td>
<td>EF143376</td>
</tr>
</tbody>
</table>

*a* Source of additional sequences: Sun and Nie (2004).

*b* Source of additional sequences: Zhao et al. (2006a).
amplification, the corresponding bases of the complete viperin gene in *Siniperca chuatsi* (Sun and Nie, 2004) were removed before alignment. The alignments of viperin gene sequences and three mtDNA segments (16S, cyt b and CO1) were conducted respectively with Clustal X 1.83 (Thompson et al., 1997) with default gap penalties. The mtDNA alignments were straightforward, while the alignment of the viperin gene showed some variations in the lengths of the sequences. The actual intron of viperin gene was located by aligning the nucleotide sequences with the published complete viperin gene sequence of *S. chuatsi* (Sun and Nie, 2004). Alignment of the 16S data was subsequently adjusted manually on the basis of secondary structure following the published model of *Galaxias brevipinnis* (Waters et al., 2000). All the first two introns of the S7 were aligned using the direct optimization method implemented in the computer program POY 3.0.11 (Gladstein and Wheeler, 2002), with parsimony as the optimality criterion. For all searches reported, we presented the best (lowest cost) trees from 100 random addition replicates (commands: -random 100 -maxtrees 3). The aligned matrix from this procedure was checked by eye, and the minor adjustments were made manually with SEAVIEW (Galtier et al., 1996). Gaps were considered as missing data rather than fifth characters, to prevent those longer than one or two bases from being taken as representing multiple events (Swofford, 1993). Pairwise distances based on the F84 model (Felsenstein, 1993) were calculated with the DAMBE program (Xia and Xie, 2001). Stationarity of nucleotide composition across taxa were examined using chi-square ($\chi^2$) tests implemented in PAUP* 4.0b10 (Swofford, 2002).

### 2.4. Phylogenetic analyses

The separate and combined dataset were used to infer phylogenies by maximum parsimony (MP) using PAUP* 4.0b10 (Swofford, 2002), maximum likelihood (ML) using PHYML 2.4.4 (Guindon and Gascuel, 2003) and recently developed partitioned Bayesian analyses (Nylander et al., 2004; Brandley et al., 2005) with MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The use of partitioned Bayesian analyses can incorporate complex mixed models of nucleotide evolution by allocating independent models to partitions of a heterogeneous dataset. This property has made it suitable for merging heterogeneous data matrices and should reduce systematic error, thus providing more accurate posterior probability estimates (e.g., Brown and Lemmon, 2007; Castoe and Parkinson, 2006; Guo and Wang, 2007). In MP analysis, a branch-and-bound search strategy was employed with all characters treated as equal weights and gaps as missing data. The best-fitting models of sequence evolution for ML and Bayesian analyses were estimated using Modeltest 3.7 (Posada and Crandall, 1998) under the Akaike information criterion (AIC; Akaike, 1974), following recent recommendations by Posada and Buckley (2004). To test for nodal reliabilities, heuristic bootstrap analyses (Felsenstein, 1985; 1000 replicates for MP and ML) were applied, with groups appearing in 50% or more of the trees in bootstrap analysis retained.

In Bayesian analyses, cyt b and CO1 was partitioned by codon position, 16S gene by the paired position and unpaired position, and the viperin gene by the individual exon and intron. The models implemented in our Bayesian phylogenetic analyses are listed in Table 3. Bayesian analyses started with randomly generated trees and four Markov chains under default heating values were run for $2 \times 10^7$ generations, with sampling at intervals of 200 generations. To ensure that the analyses were not trapped on local optima, the dataset was run three times independently. The length of the burn-in was determined by plotting the ln likelihood of the trees sampled, by examining the potential scale reduction factor for all variables, and plotting the frequency of bifurcations in the sampled trees in the two runs using AWFTY (Wilgenbusch et al., 2004). After determining chain convergence, which generally occurred within the first 2–5 million generations of each analysis, we followed a conservative approach by discarding all samples obtained during the first 8 million generations as “burn-in”. We then generated 50% majority-rule consensus trees with posterior probability values for each node in the context of the final 12 million generations obtained during each analysis.

Partitioned Bremer support analysis (PBS; Bremer, 1994), as calculated using PAUP* 4.0b10 (Swofford, 2002) and TreeRot V2 (Sorenson, 1999), was used to assess the respective contribution of each gene to the total nodal Bremer support. We also tested for incongruent phylogenetic signal between nuclear and mt genes using the partition homogeneity (PHT) test (essentially the incongruence length difference (ILD) test of Farris et al. (1994) in PAUP* 4.0b10 (Swofford, 2002). Even if incongruence among datasets was

Table 2

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer sequence</th>
<th>Approximate product length (bp)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1f</td>
<td>5′-GGG AAC AGA GCA GCT GTG ATT C-3′</td>
<td>1039</td>
<td>This study</td>
</tr>
<tr>
<td>G1r</td>
<td>5′-AAG GTG TTG ATG ACC GAG TTG ATC-3′</td>
<td>1039</td>
<td>This study</td>
</tr>
<tr>
<td>G2f</td>
<td>5′-TCA ACT CGG TCA TCA ACA CTT TCA-3′</td>
<td>1720</td>
<td>This study</td>
</tr>
<tr>
<td>G2r</td>
<td>5′-TCC ACA CAT ATT TCC CTC CTC TC-3′</td>
<td>1720</td>
<td>This study</td>
</tr>
<tr>
<td>G3f</td>
<td>5′-ATG CCT TCT CTG GAC TGT CG-3′</td>
<td>377</td>
<td>This study</td>
</tr>
<tr>
<td>G3r</td>
<td>5′-CTA TCT ATC TAT CTA TCT ATC TAT-3′</td>
<td>377</td>
<td>This study</td>
</tr>
</tbody>
</table>

Fig. 1. Diagram of *Siniperca chuatsi* viperin gene structure drawn according to Sun and Nie (2004) and the primer positions in this study. The gene diagram is drawn approximately to scale, except for the arrows designating the primer localization and direction of PCR. Black boxes represent exons.
observed, the combined data approach was adopted, which has been favored by many to fully utilize genetic information in deriving phylogenetic estimates (Rokas and Carroll, 2006).

### 2.5. Alternative phylogenetic hypotheses tests

To compare competing siniperids topologies, sitewise log-likelihoods were calculated for each topology in PAUP* and used as input for CONSEL 0.1f program (Shimodaira and Hasegawa, 2001). CONSEL was used to calculate the probability values according to the approximately unbiased (AU) test using the multiscale bootstrap approach to remove this bias (Strimmer and Rambaut, 2002). Three mt genes (16S, cyt b and CO1) were combined and analyzed simultaneously (~2312 bp). The topologies obtained by various analytic methods were similar (MP tree length = 1680, Cl = 0.6232, and RC = 0.3058) and exhibited much improved resolution and nodal support than either mtDNA gene alone (see Table 5).

### 3. Results

#### 3.1. Sequence characteristics

Summaries of sequence characteristics for individual and combined gene partitions are given in Table 4 from 7307 characters of sequence data for 13 taxa. The concatenated mtDNA gene fragments consisted of 2312 sites, with 733 variable characters (VC) and 489 parsimony-informative characters (PIC). The concatenated nuclear datasets consisted of a matrix of 4995 base pairs (bp). Of these, 1737 bp were VC and 667 bp were PIC. The aligned nuclear dataset was assembled from three fragments: Viperin gene dataset (3502 bp/939 VC/300 PIC); S7 intron 1 (851 bp/391 VC/224 PIC); S7 intron 2 (642 bp/407 VC/143 PIC). Viperin gene showed the secondarily lowest proportion of PIC (8.57%) compared with the other gene sequences. Base composition was AT-rich biased in all the six genes. The viperin gene was the secondly most AT-rich (59.02%), and showed the lowest transition/transversion ratio (maximum likelihood Ti/Tv = 1.09), contrasted with obvious tendency of transitions in the mtDNA genes (4.57 in cyt b, 6.23 in CO1 and 7.68 in 16S). F84 distances within the ingroup taxa ranged from 0.7 to 43.2% for S7 intron 1 (19.5% on average), from 0.8 to 74.2% for S7 intron 2 (26.8% on average), from 0.3–12.5% for viperin (6% on average), and from 0.5–20.1% for combined mt DNA sequences (11.1% on average).

As shown in Table 5, we were unable to obtain sequences from intron 3 to exon 6 of viperin gene in C. herzi, from exon 1 to intron 1 in N. spinosus and from intron 4 to exon 6 in L. maculatus, and therefore the three species have fewer characters than all other species in the matrix. The effects of such incomplete matrices are difficult to predict, but simulations suggest the inclusion of a limited amount of missing data was unlikely to distort the phylogenetic results (Wiens, 2003). The viperin gene fragment obtained ranges in size from 979 to 3181 bp. The detailed length of each exon and intron in viperin gene sequences across all taxa was listed in Table 5. For example, the first intron varies from 79 (Coreoperca whiteheadi) to 104 bp (L. maculatus) in size, but within the genus Siniperca, the size is constant (86 bp).

### 3.2. Phylogenetic analyses of different genes and gene combinations

#### 3.2.1. Mitochondrial genes

Three mt genes (16S, cyt b and CO1) were combined and analyzed simultaneously (~2312 bp). The topologies obtained by various analytic methods were similar (MP tree length = 1680, Cl = 0.6232, and RC = 0.3058) and exhibited much improved resolution and nodal support than either mtDNA gene alone (see Table 5).
Appendix 1–3). MP and ML analyses of the combined mt data were shown respectively in Appendix 4A and 5A, and Bayesian tree was presented in Fig. 2A. As expected, the monophyly of both Siniperca and Coreoperca was recovered with strong posterior probability (PP = 1.0 and 1.0, respectively). Similar support values for the corresponding nodes were also obtained by MP (100% and 90%, respectively) and ML (100% and 90%, respectively). However, the monophyly of all recognized sinipercids, i.e., the family Siniperidae, was not corroborated in all analyses, with *N. spinosus* joining Siniperca in Bayesian tree (PP = 0.97) or Coreoperca in MP tree (BP = 57). The interrelationships of Siniperca were well-resolved and strongly supported for almost all nodes. The basal split in the genus is between *Siniperca roulei* and a strongly supported subclade of other *Siniperca* species (PP = 1.0, BP = 85% and 93% for MP and ML, respectively). Within the remaining *Siniperca* species, *S. chuatsi* and *S. kneri* consistently showed a robust sister species and *Siniperca obscura* was identified as the sister taxon to them. Similarly, *Siniperca fortis* and *Siniperca undulata* consistently showed a sister species and *Siniperca scherzeri* was identified as the sister taxon to them, despite the MP bootstrap value lower than 50%. In the genus *Coreoperca*, *C. herzi* was moderately placed as the sister taxon to *Coreoperca kawamebari* and *C. whiteheadi* under the

### Table 5

The length of each exon and intron in viperin gene sequences across all taxa.

<table>
<thead>
<tr>
<th>Species</th>
<th>Exon 1</th>
<th>Intron 1</th>
<th>Exon 2</th>
<th>Intron 2</th>
<th>Exon 3</th>
<th>Intron 3</th>
<th>Exon 4</th>
<th>Intron 4</th>
<th>Exon 5</th>
<th>Intron 5</th>
<th>Exon 6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. chuatsi</em></td>
<td>358*</td>
<td>86</td>
<td>162</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>865</td>
<td>33</td>
<td>362</td>
<td>373*</td>
<td>2972</td>
<td></td>
</tr>
<tr>
<td><em>S. scherzeri 1</em></td>
<td>358*</td>
<td>86</td>
<td>162</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>891</td>
<td>33</td>
<td>368</td>
<td>374*</td>
<td>2982</td>
<td></td>
</tr>
<tr>
<td><em>S. scherzeri 2</em></td>
<td>358*</td>
<td>86</td>
<td>162</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>895</td>
<td>33</td>
<td>368</td>
<td>374*</td>
<td>3020</td>
<td></td>
</tr>
<tr>
<td><em>S. kneri</em></td>
<td>358*</td>
<td>86</td>
<td>162</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>874</td>
<td>33</td>
<td>363</td>
<td>374*</td>
<td>2985</td>
<td></td>
</tr>
<tr>
<td><em>S. fortis</em></td>
<td>358*</td>
<td>86</td>
<td>162</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>888</td>
<td>33</td>
<td>374</td>
<td>374*</td>
<td>3005</td>
<td></td>
</tr>
<tr>
<td><em>S. obscura</em></td>
<td>358*</td>
<td>86</td>
<td>162</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>881</td>
<td>33</td>
<td>374</td>
<td>374*</td>
<td>3006</td>
<td></td>
</tr>
<tr>
<td><em>S. undulata</em></td>
<td>358*</td>
<td>86</td>
<td>162</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>810</td>
<td>33</td>
<td>372</td>
<td>374*</td>
<td>2988</td>
<td></td>
</tr>
<tr>
<td><em>S. roulei</em></td>
<td>358*</td>
<td>86</td>
<td>162</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>870</td>
<td>33</td>
<td>372</td>
<td>374*</td>
<td>2988</td>
<td></td>
</tr>
<tr>
<td><em>C. herzi</em></td>
<td>358*</td>
<td>81</td>
<td>162</td>
<td>206</td>
<td>172*</td>
<td>–</td>
<td>––</td>
<td>––</td>
<td>––</td>
<td>9</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td><em>C. kawamebari</em></td>
<td>358*</td>
<td>81</td>
<td>162</td>
<td>186</td>
<td>230</td>
<td>94</td>
<td>150</td>
<td>795</td>
<td>33</td>
<td>382</td>
<td>208*</td>
<td></td>
</tr>
<tr>
<td><em>C. whiteheadi</em></td>
<td>358*</td>
<td>79</td>
<td>162</td>
<td>194</td>
<td>230</td>
<td>99</td>
<td>150</td>
<td>799</td>
<td>33</td>
<td>704</td>
<td>373*</td>
<td></td>
</tr>
<tr>
<td><em>N. spinosus</em></td>
<td>–</td>
<td>–</td>
<td>8*</td>
<td>3*</td>
<td>177</td>
<td>230</td>
<td>89</td>
<td>150</td>
<td>810</td>
<td>33</td>
<td>155</td>
<td></td>
</tr>
<tr>
<td><em>L. maculatus</em></td>
<td>349*</td>
<td>104</td>
<td>162</td>
<td>178</td>
<td>233</td>
<td>101</td>
<td>156</td>
<td>402*</td>
<td>–</td>
<td>–</td>
<td>1685</td>
<td></td>
</tr>
</tbody>
</table>

---

Fig. 2. A majority-rule consensus of trees sampled from the posterior distribution (at stationarity) of the Bayesian analysis, with posterior probabilities above the branch. Numbers behind the nodes are arbitrarily defined numbers represent the clade that follow and which can be used in Table 6.
3.2.2. Nuclear genes

MP consensus trees based on analyses of individual viperin dataset and combined nuclear dataset (viperin + S7 intron 1 and intron 2; \( \approx 4995 \) bp) were shown in Appendix 4B and C, respectively. The topologies of two genealogical trees contrasted in several aspects, especially for the placement of S. undulata and S. fortes. Moreover, the 50% majority consensus trees from the ML and Bayesian analyses (Appendix 5B and C, Fig. 2B and C) were not completely consistent with the corresponding MP tree. Notably, the topological differences from different methodologies were weakly supported (BP < 50%, PP < 0.9). The combined nuclear data tree tended to have overall greater resolution and higher nodal supports. Combined nuclear data analyses consistently identified that the basal split in the genus Siniperca is between S. undulata and a strongly supported subclade of the remaining Siniperca species (PP = 1.0, BP = 85 and 93% for MP and ML, respectively). Within the latter subclade, S. fortes showed to be the sister taxon to other Siniperca species except S. undulata (MP, 68%; ML, 89%; Bayesian, 1.0). As with the mt DNA analyses, our combined nuclear data analyses clearly confirmed closet affinity between S. chuatsi and S. kneri. Contrast to mtDNA, our nuclear data analyses strongly supported that C. herzi and C. kawamebari formed as sister species, being sister to C. whiteheadi. As shown in Fig. 2B and C, Appendix 4B and C, and Appendix 5B and C, the reciprocal monophyly of the genus Siniperca and Coreoperca as well as the family Siniperidae was all recovered in viperin and combined nuclear analyses.

3.2.3. Combined data (mt DNA + nuclear DNA)

 Parsimony analysis using equal weights only resulted in a single MP tree (tree length = 4046, CI = 0.7731, and RC = 0.6183), which is shown in Appendix 4D. ML and Bayesian analyses, identical topology with each other, were essentially consistent with the MP tree except for the placement of S. undulata and S. fortes. Clearly, comparisons between the MP and Bayesian trees suggested that the Bayesian tree provides higher resolution than the MP. As well with combined nuclear data, our combined data (mtDNA + nuclear DNA) analysis clearly confirmed a well-supported monophyletic sinipericid fish with two distinct genera in accordance with Liu and Chen (1994), one represented by Siniperca and the other by Coreoperca. ML analysis of the combined data is presented in Appendix 5D. This multiple gene tree is completely resolved and strongly supported for all nodes except for the positions of S. roulei (69%) and S. obscura (53%). Similar support values for the two nodes were also obtained by Bayesian analyses (Fig. 2D; 0.84 for S. roulei, 0.65 for S. obscura). In the Siniperca, S. undulata occupied the most basal position, followed by S. fortes, S. scherzeri, S. roulei, S. obscura, and last two most recently diverged sister species S. chuatsi and S. kneri. As for the interrelationships in Coreoperca, they are identical with those of nuclear datasets.

The partition homogeneity tests showed that no heterogeneity in phylogenetic signal was detected between any of the mtDNA genes, and that partitions were significantly heterogeneous between viperin and S7 intron 1 and 2 (\( P = 0.013 \)), and between combined mtDNA and combined nuclear DNA (\( P = 0.001 \)), etc. Partitioned Bremer values (Table 6) reveals that the majority of phylogenetic information from the combined gene-based topologies was contributed by nuclear character, with S7 intron 1 holding the highest percentage (59.35%), followed by viperin (27.43%), and S7 intron 2 (10.22%), in sharp contrast to the mitochondrial characters (2.99% in total). The relatively weak influence of mitochondrial partition upon analysis of complete set of genes was somewhat surprising, given their relatively high degree of genetic divergence and rich informative characters. The breakdown of the PBS values indicated some conflict between the mt DNA datasets and the nuclear DNA datasets, because 6 of 10 nodes resolved in our combined dataset consensus tree had conflicting PBS values. This pattern may have resulted from either differences in internal homoplasly within each partition or potential conflicting signals, as indicated by PHT test, between mtDNA and nuclear DNA.

The competing hypotheses inferred using different partition datasets were also statistically tested among each other by AU, SH and Templeton tests (Table 7). Under the combined dataset, the three tests supported that S7 intron 2 Bayesian tree and mtDNA Bayesian topologies including those from combined mtDNA, 16S, COI and cyt b are all safely rejected (\( P < 0.05 \)). However, the hypotheses from viperin and S7 intron 1 can only be significantly rejected by AU test (\( P = 0.029, < 0.001 \)) and Templeton test (\( P = 0.0061, 0.0005 \)), but cannot by SH test (\( P = 0.611, 0.076 \)). This echoes Strimmer and Rambaut (2002) in that the SH test is more conservative (i.e. less likely to reject alternative topologies under consideration) than AU test. The combined dataset Bayesian hypothesis and monophyly constraint hypothesis were not significantly different, which were covered similar log-likelihood score and tree length. Thus, in our view, the hypotheses based on ML and Bayesian analyses of combined datasets (Appendix 5D and Fig. 2D) represent the best current estimate of sinipericids phylogeny in the present phylogenetic reconstruction.

4. Discussion

4.1. Monophyly and interspecific relationships of sinipericid fish

Traditionally, sinipericid fish are composed of twelve recognized extant species. With the unusual morphology and relatively few perciform fish in freshwater, these fish have been the subject of

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Mitochondrial datasets</th>
<th>Nuclear datasets</th>
<th>Combined nuclear and mt gene trees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16S</td>
<td>COI</td>
<td>Cyt b</td>
</tr>
<tr>
<td>1</td>
<td>−1</td>
<td>5</td>
<td>−13</td>
</tr>
<tr>
<td>2</td>
<td>−2</td>
<td>−2</td>
<td>−4</td>
</tr>
<tr>
<td>3</td>
<td>−2</td>
<td>−2</td>
<td>−4</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>−2</td>
<td>−2</td>
<td>−4</td>
</tr>
<tr>
<td>6</td>
<td>−2</td>
<td>−2</td>
<td>−4</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>−16</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>7</td>
<td>−4</td>
</tr>
<tr>
<td>10</td>
<td>−3</td>
<td>0</td>
<td>−1</td>
</tr>
<tr>
<td>Total PBS</td>
<td>−9</td>
<td>23</td>
<td>(5.74%)</td>
</tr>
</tbody>
</table>

* Nodes are numbered as the combined data Bayesian tree in Fig. 2D.
several taxonomic and phylogenetic studies for decades. However, as summarized by Chen et al. (2007), different views still exist over the taxonomic status and phylogeny of siniperids. In support of Zhou et al. (1988), Kong and Zhou (1993) suggested that the siniperid fish is a monophyletic group consisting of two genera, namely, Siniperca and Coreoperca. Zhou et al. (1988), Kong and Zhou (1993) and Liu and Chen (1994) confirmed that the siniperid fish is a monophyletic group consisting of two genera, namely, Siniperca and Coreoperca, and the genus Coreoperca was unwarranted and Coreosiniperca roulei is a member of the genus Siniperca. The interspecific relationships within this group have been also equivocal probably due to rapid cladogenesis resulting in short internodes that have not been resolved because of insufficient sequence data used in previous studies (e.g., Chen et al., 2007; Zhao et al., 2006a,b).

In the present study, although individual genes examined failed to provide a well-supported phylogeny, the ML and Bayesian analyses of concatenated mt and nuclear data, consistently produced a robust, well-resolved tree. Our phylogenies strongly upheld the monophyly of siniperid fish (Sinipercaeidae), Siniperca and Coreoperca, which agrees well with several key features of recent perspectives of siniperid systematics. Coreosiniperca roulei (= Siniperca roulei) does not form an independent group, instead merges into the genus Siniperca, which is in accord with the results of Liu and Chen (1994) and Zhao et al. (2006a,b). The internal affinities within Coreoperca are shown to be novel, and our results differ from all previous hypotheses (Fig. 3). Our analyses clearly indicate that C. whiteheadi is the sister taxon to C. herzi plus C. kawamebari. Similarly, the interrelationships in the genus Siniperca are different from previous hypotheses (Fig. 3). Our results demonstrate that S. undulata is the sister taxon to the other members of Siniperca, within the clade containing the other members of the genus, S. chuatsi and S. kneri are sister species, next joined by S. obscura, S. roulei, S. scherzeri and finally by S. fortis.

Several interesting phenomena with respect to the phylogenetic relationships of siniperid fish are observed in the present study. Firstly, mitochondrial gene whenever single or simultaneous analysis cannot recover the monophyly of siniperids, while the nuclear vipherin gene, the S7 intron 2 and the total data supported the monophyly of this family with higher bootstrap value or posterior possibility. Secondly, the monophyly of two genera Siniperca and Coreoperca were supported by vipherin gene, simultaneous mtDNA or nuclear DNA and combined data in three analytical methods. Thirdly, most single gene and the total data supported the mono-

![Fig. 3. Competing hypotheses of phylogenetic relationships within siniperid fish deduced from (A) morphology (Liu and Chen, 1994). (B) Complete cyt b gene sequence (Chen et al., 2007).](image-url)
phyly of genus Siniperca, while single cyt b and 16S cannot corroborate the genus Coreoperca as a monophyletic group.

Taking all evidence together, we can find that nearly no two analyses have come to a completely identical result regarding affinities among these seven species in Siniperca (with the exception of ML and Bayesian trees from combined mt and nuclear DNA dataset in the present study), with conclusions varying depending on the analytic methods and character type used. However, given the large size of our dataset and the robust support of many nodes, we consider our total data gene tree as the preferred interpretation of sinipercid fish relationships. AU, SH and Templeton tests suggest that the combined data topology we recovered (Fig. 2D and Appendix 5D) were most possible estimate for present dataset, although the results from combined nuclear DNA cannot be statistically rejected by all the tests at 5% significant level.

4.2. Molecular evolution patterns and phylogenetic utility of different mt and nuclear genes in sinipercid fish

Our results indicate that nuclear genes/introns have much advantage over mitochondrial genes in equal weights parsimony analysis. Nuclear genes/introns had universally higher values of CI as compared to mitochondrial genes (Table 4), and generally provided more in the way of partitioned Bremer support than CI as compared to mitochondrial genes (Table 6). These results suggest the view provided more in the way of partitioned Bremer support than CI as compared to mitochondrial genes (Table 6). These results suggest the view.

Given the detected conflict between the mitochondrial genes and nuclear DNA, it is important to identify the source of this disagreement. By performing Bayesian analysis using a GTR + site-specific rates (SSR) model with rate categories corresponding to gene, it was possible to quantitatively compare the substitution patterns presented by different mitochondrial genes and nuclear genes/introns. Simon et al. (1994) concluded in general terms that nuclear genes evolve more slowly than mitochondrial genes, making nuclear genes better markers for deep divergence. This may be the general tendency for exons, but it does not necessarily apply when comparing nuclear introns with mitochondrial genes. When we consider our data alone this pattern is not seen because the relative rate presented by S7 intron 1 (0.905) and intron 2 (1.432) was greater than 16S (0.327). Indeed, this pattern could already be expected because the mean F84 distance values obtained for S7 intron 1 (0.195) and intron 2 (0.268) was much greater than that for 16S.

It was previously seen that the nuclear DNA genes/introns showed greater base compositional bias than the mitochondrial genes, but other patterns of nucleotide substitution, such as the Q matrix of transformation, may also be important. In general, the instantaneous rate matrices for mitochondrial genes was more asymmetrical relative to those presented by nuclear DNA (viperin, S7 intron 1 and intron 2) and were also more skewed towards one type of change over another (Fig. 4B), although in all cases there was a higher overall rate of transitions. The obvious consequence of these highly skewed transformation rate matrices is greater levels of homoplasmy, which are not easily corrected for.

Another parameter that often differs between nuclear intron and mitochondrial gene is the shape of the gamma distribution, as given by \( \alpha \) value, describing the among-site rate variation. As can be seen from Fig. 4C, there is much more heterogeneity in among-site rate variation in mitochondrial genes than in nuclear S7 intron 1 and intron 2, with the most heterogeneous in cyt b gene, which had the lowest \( \alpha \) value. However, all the nuclear sequences presented lower proportion of invariable sites (Pi values), which is contrast to that of Lin and Danforth (2004), who demonstrated that the proportion of invariable sites is positively correlated with \( \alpha \) value. These authors also showed a positive correlation between \( \alpha \) value and CI, the consistency index, which suggests that data partitions with more heterogeneous substitution rates show a higher level of homoplasy. This correlation was detected by us, with the nuclear datasets having higher CI values (viperin, 0.91; S7 intron 1, 0.86; S7 intron 2, 0.9) than the mt genes (cyt b, 0.62; 16S, 0.71; CO1, 0.62).

The three mt genes included in this study failed to recover the monophyly of sinipercid fish, thus making the combination of all mt genes unwarranted for providing persuasive evidence of the resolution of all parts of the trees. S7 intron 1 and intron 2 datasets in present study proved especially informative in phylogenetic reconstruction of the sinipercid fish. Both introns evolved at a rapid rate and no sign of significant saturation were observed. The S7 intron 1 was found to contribute the most phylogenetic signal while S7 intron 2, only less than S7 intron 1 and viperin to the final tree (Fig. 2D; see Table 6). However, like the mitochondrial genes in this study and cyt b in Chen et al. (2007), S7 intron 1 failed to confirm the monophyly of sinipercid fish either.

It is notable that our study provides valuable information on the utility of nuclear viperin gene in the phylogenetic reconstruction for the first time. The viperin gene was explored in the present study to reconstruct phylogeny of sinipercid fish, encouraged by
the fact that this gene has been shown to contain valuable phylogenetic signals corroborating the monophyly of Siniperca, Coreoperca, and a single origin of sinipercid fish. However, the vipherin gene alone cannot resolve the interspecific relationships of the genus Siniperca. It is expected that this novel nuclear marker would be better suited for resolving suprageneric (e.g., interfamilial) relationships among periform and other fish.

In conclusion, our results demonstrate that using just one of these markers would not have provided a well-resolved phylogeny, even with the most sophisticated analytical strategies we could find in the literature. This lack of resolution is largely due to insufficient phylogenetic information in individual loci. Analysis of these heterogeneous data using partitioned models of sequence evolution removes the problems associated with combining data and appears to be the most logical way to analyze heterogeneous data. By further combining mt DNA and nuclear datasets, we have been able to provide a well-supported, completely sampled phylogenetic hypothesis for sinipercid fish.

Acknowledgments

We are grateful to Dr T. King, L. Yang, Y. Zhang, Z. Li, A. Li, D. Wang, L. Zhou and Mrs. M. Pan, for assistance in collecting specimens. We also thank two anonymous reviewers for their insightful comments on this manuscript. This study was supported by a joint fund from the National Natural Science Foundation of China and the Government of Guangdong Province (U0631010). Dali Chen was supported by the National Natural Science Foundation of China (30800094) and Education Foundation for Young Teachers of Sichuan University (07056). Xianguang Guo was supported by the National Natural Science Foundation of China (30700062), and the Knowledge Innovation Program of the Chinese Academy of Sciences (CIB-2007-IYQ-01, KSCX2-YW-Z-005).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.01.039.

References