Genetic Diversity and Phylogeography of *Taenioides cirratus* in Five Geographical Populations Based on Mitochondrial COI and Cytb Gene Sequences

Guoqing Zhang, Cheng Chen, Wenxuan Lu, Jing Li, Ting Fang, Kun Yang, Xiuxia Zhao, Na Gao, and Yangyang Liang

Key Laboratory of Freshwater Aquaculture and Enhancement of Anhui Province, Fisheries Research Institute, Anhui Academy of Agricultural Sciences, Hefei, China

Correspondence should be addressed to Yangyang Liang; liangyy10214@126.com

Received 15 November 2022; Revised 10 June 2023; Accepted 15 June 2023; Published 28 June 2023

Academic Editor: Osman Sabri Kesbi

The genetic diversity of *Taenioides cirratus* was investigated on the basis of mitochondrial cytochrome c oxidase subunit I (COI) and cytochrome b (Cytb) gene sequences. A total of 159 specimens collected from the Chaohu Lake (CL), Nansihu Lake (NL), Taihu Lake (TL), Pearl River (PR), and Nandu River (NR) were sequenced. The total length of the sequence was 2485 bp with 412 polymorphic sites. A total of 73 haplotypes were identified, with Hap1 being the most widely distributed. The PR and NR populations showed high genetic diversity, while the CL population showed low genetic diversity. TL and NL showed high haplotype diversity but low nucleotide diversity. The analysis of molecular variance demonstrated that the sequence variations were mainly occurred among populations. *T. cirratus* populations are declining, and rare alleles are present at low frequencies, as analysed using a neutral test and a mismatched distribution analysis. There was a relatively high level of genetic differentiation among the populations of the Yangtze River basin (including NL, PR, and NR ($F_{ST} > 0.15$)). The two similar phylogenetic trees constructed by the maximum likelihood (ML) and Bayesian inference (BI) methods presented three major lineages, of which lineage II contains haplotypes from PR and NR, lineage III contains haplotypes from CL, NL, TL, and PR, whereas lineage I contains only a portion of haplotypes from NR. Based on the neutral test, mismatch analysis, and Bayesian Skyline Plot (BSP), geological and climatic events were inferred to have played an important role in the historical dynamics of *T. cirratus* population. Hap1, Hap25, and Hap58 were inferred as possible ancestral haplotypes by network analysis. Our study offers an essential foundation for resource preservation and additional taxonomic clarification of *T. cirratus*.

1. Introduction

Gobiidae is a warm-water carnivorous small fish family that occurs along the world’s coasts, except in the Arctic and Antarctic. *Taenioides cirratus*, which belongs to Amblyopinae, Gobiidae, and Gobioidae, is native to estuaries along the offshore coast of the East China Sea and South China Sea [1]. In recent years, it has invaded many inland lakes, such as the Chaohu Lake [2], Nansihu Lake [3], and Gaoyou Lake [4]. Liang et al. [5] speculated that under future climate change scenarios, the suitable distribution area for *T. cirratus* will greatly increase. Additionally, Amblyopinae taxonomy is the subject of debate [6], particularly the proposition to divide the geographical populations from Hainan Province (corresponds to the NR in present study), the Pearl River, and the Yangtze River into three independent effective species [7].

Over the past few years, genetic sequencing has been widely used in the analysis of the population structure of fishery resources, which is difficult to distinguish based on behaviour and morphology [8]. It is useful for the investigation of species identification, population structure, and genetic differentiation of fluvial fisheries resources [9–11]. Abundant data demonstrate that sequence data for...
the mitochondrial cytochrome c oxidase subunit I (COI) and cytochrome b (Cytb) genes are often suitable for individual-based species/specimen identification and very useful for phylogenetic reconstructions [12–14], including of numerous fish taxa [5, 15].

To assess the risk of extinction for many species, the IUCN adopted the Red List categories of plants and animals. *T. cirratus* has been listed as “Data Deficient” in the IUCN Red List of Threatened Species in 2011 [16]. This study used molecular data from mitochondrial COI and Cytb gene sequences to explore the genetic diversity and population structure of *T. cirratus*. It provides new insights for the debate on the taxonomic affiliation of *T. cirratus* and a rationale to further clarify the conservation of *T. cirratus* resources as well as guidance for the prevention and control of *T. cirratus* invasion.

### 2. Materials and Methods

#### 2.1. Experimental Materials.

From June 2017 to August 2020, a total of 159 specimens of *T. cirratus* were caught by baited benthic fyke nets from the Chaohu Lake (CL), Taihu Lake (TL), Nansihu Lake (NL), Pearl River (PR) estuary, and Nandu River (NR) estuary (Figure 1). Although NL is located in the Huaihe River basin, it is connected with the Yangtze River basin by the South-to-North Water Diversion, so we grouped it into the Yangtze River basin in this study. After catching, the muscle tissue of each specimen was collected and preserved in 95% ethanol.

#### 2.2. DNA Extraction.

Total genomic DNA of each specimen was extracted using the standard phenol-chloroform method [17]. Muscle tissue was cut into pieces, and 500 μL HOM buffer and 10–15 μL (10 mg/mL) protease K (Sigma–Aldrich) were added and digested at 55°C for more than 3 hours. Then, 500 μL NaCl (4.5 mol/L) and 300 μL trichloromethane were added and centrifuged at 13,000 rpm for 10 minutes. The supernatant (approximately 800 μL) was mixed with about 560 μL isopropl alcohol (0.7 times the supernatant volume) and placed at -20°C for at least 1 hour. The solution was centrifuged at 13,000 rpm for 10 minutes to get the DNA sediment.

Based on the complete mitochondrial genome of *T. cirratus*, specific primers were designed according to the conserved sequences of the COI (F-5′-ACATTGGGACCC-TTTATC-3′, R-5′-TTGCCGTTAGTTAACCAGATG-3′) and Cytb (F-5′-CATCATTCCGAGGCCGT-3′, R-5′-GGCTTACAAGACCGGC-3′) genes. The total PCR volume was 50 μL, containing 5 μL of 10 × PCR buffer, 4 μL of dNTP (2.5 mmol/L), 2 μL primers (10 mmol/L), 0.8 μL of Taq DNA polymerase, 1 μL of template DNA, and 37.2 μL ddH₂O. The samples were amplified by a PCR apparatus, and the reaction procedure was strand denaturation at 94°C for 4 minutes, followed by denaturation at 94°C for 45 seconds, annealing at 53°C for 45 seconds, extension at 72°C for 1 minute, for 30 cycles, and a final extension at 72°C for 10 minutes. PCR products were sent to a commercial company for bidirectional sequencing [18]. The COI and Cytb gene sequences obtained for the studied *T. cirratus* populations are deposited in GenBank (accession numbers OM849673–OM849724 for Cytb and OM849725–OM849774 for COI).

#### 2.3. Sequence Analysis.

Sequences were aligned using MEGA 11 [19]. Haplotype, variable sites, and mismatch distribution were counted using DNASP v6 [20]. Statistics for every population were calculated and compared among populations using Arlequin 3.5.1 [21], and analysis of molecular variance (AMOVA) was used to estimate the source of genetic variation.

Phylogenetic relationship analysis was conducted using both maximum likelihood (ML) and Bayesian inference (BI) methods, with *Odontobutis potamophila* as the outgroup. BI tree was constructed using MrBayes v3.2.3 [22], four Markov chains (Monte-Carlo Markov Chains (MCMC)) were run simultaneously for 200,000,000 generations, and one tree was randomly selected every 1,000 generations using Tracer 1.6 [23] for convergence diagnostics and determination of discard values. The burn-in value was determined, and the top 20,000 trees were discarded based on the diagnostic results (standard deviation of divergence frequency <0.05). The support of nodes was determined by posterior probability. ML tree was constructed using PhyML 3.0 [24], and the support of nodes was calculated using the nonparametric bootstrap method [25] and repeated 1,000 times. The generated phylogenetic tree was visualised with FigTree v1.4.2 [26]. In addition, phylogenetic network diagrams of haplotypes were constructed using POPART software [27]. For each dataset, Bayesian Skyline Plot (BSP) analyses were implemented in BEAST 2.7.3 [28] using a strict clock with *T. cirratus* mass/gene mutation rates (2.20%/Myr) and the appropriate substitution model from PhyloSuite v1.2.3 [29, 30]. MCMC runs of up to 100 million steps yielded effective sample sizes (ESSs) of at least 200.

### 3. Results

#### 3.1. Haplotype and Nucleotide Diversity.

A total of 159 individuals were sequenced for the COI and Cytb genes (2485 bp). A total of 412 polymorphic sites were detected, yielding 73 haplotypes. The haplotypes were unevenly distributed among the populations. Hap1 (26.88%) and Hap25 (1.88%) were shared among different populations, and the other 71 haplotypes were singletons. The PR and NR populations showed high haplotype diversity (h) and nucleotide diversity (π), the TL and NL showed high h but low π, and the CL showed low h and π (Table 1).

#### 3.2. Neutral Test.

According to haplotype mismatch distribution analysis, Tajima’s D value for the CL population was zero, indicating a stable population, while Tajima’s D value for the TL and NL populations was negative with a p < 0.01, reaching the significance level, indicating that the populations may have expanded and enriched rare alleles. Overall, Tajima’s D value for the *T. cirratus* population was greater than zero, and the p < 0.01 reached the significance
level, indicating that the population is shrinking and loci are subject to balanced selection (Table 2).

3.3. Fst Analysis. The pairwise Fst distance varies from 0.09697 to 0.79203. Lower pairwise Fst distances were observed in the CL, TL, and NL populations, ranging from 0.09697 to 0.14745. Overall, the maximum genetic distance was observed between specimens from NR and other populations. All comparisons reached significance (p < 0.05) (Table 3).

3.4. Analysis of Molecular Variance. Analysis of molecular variance resulted in which the variance among populations within groups (72.02%) was always higher than that within populations (27.98%) (Table 4).

3.5. Haplotype Network. The median-joining haplotype network of T. cirratus is shown in Figure 2. Structure I mainly consists of haplotypes from CL, NL, TL (substructure II), and PR (substructure I and III), and Hap1 is located in the centre of structure I. Structure II mainly consists of haplotypes from PR and NR, and Hap25 is located in the centre of structure II. Structure III consists of some haplotypes from NR, and Hap58 is located in the centre of structure III.

3.6. Phylogenetic Analysis. Under two different analytical strategies, we found almost congruent phylogenies. The phylogenetic trees constructed using the BI and ML methods displayed congruent topologies. Phylogenetic analysis identified three distinct lineages of T. cirratus. Some NR populations clustered into lineage I, lineage II stands for populations from PR and NR, and lineage III stands for populations from CL, NL, TL, and PR. A few specimens from the PR and NR populations showed no clustering pattern (Figures 3 and 4).

3.7. Mismatch Distribution and BSP. The historical demographic dynamics of T. cirratus were inferred from mismatch distribution. Mismatch distribution showed that both lineage I and lineage II were identified multi-modals, whereas lineage III was identified unimodal (Figure 5). BSP of the effective population size of different

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample size</th>
<th>Number of haplotypes</th>
<th>Haplotype distribution</th>
<th>Haplotype diversity (h)</th>
<th>Nucleotide diversity (π)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>25</td>
<td>1</td>
<td>Hap1</td>
<td>0.000 ± 0.000</td>
<td>0.00000 ± 0.0000</td>
</tr>
<tr>
<td>TL</td>
<td>23</td>
<td>15</td>
<td>Hap1~15</td>
<td>0.912 ± 0.024</td>
<td>0.00607 ± 0.01292</td>
</tr>
<tr>
<td>NL</td>
<td>25</td>
<td>6</td>
<td>Hap1, Hap16~20</td>
<td>0.633 ± 0.050</td>
<td>0.00054 ± 0.00108</td>
</tr>
<tr>
<td>PR</td>
<td>30</td>
<td>20</td>
<td>Hap21~40</td>
<td>0.940 ± 0.032</td>
<td>0.03289 ± 0.00957</td>
</tr>
<tr>
<td>NR</td>
<td>56</td>
<td>34</td>
<td>Hap25, Hap41~73</td>
<td>0.925 ± 0.027</td>
<td>0.03376 ± 0.00423</td>
</tr>
<tr>
<td>Total</td>
<td>159</td>
<td>73</td>
<td>Hap1~73</td>
<td>0.914 ± 0.018</td>
<td>0.05757 ± 0.00290</td>
</tr>
</tbody>
</table>

Figure 1: Sampling sites of T. cirratus.
lineage revealed that lineage I showed rapid population growth after 13 kya B.P., lineage II experienced a population bottleneck effect between 26 and 44 kya B.P., whereas lineage III displayed a slight population decline after 32 kya B.P. (Figure 6).

4. Discussion

4.1. Genetic Diversity. As an important basis of biodiversity, genetic diversity is the result of long-term survival, evolution, and adaptation of species [31, 32]. Two important
indicators to measure genetic diversity are haplotype diversity and nucleotide diversity. The populations of *T. cirratus* as a whole were characterized by low nucleotide diversity ($\pi = 0.058$), accompanied by high haplotype diversity ($h = 0.914$). Analysis of molecular variance showed that the variance among populations (72.02%) was always

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{BI tree of *T. cirratus* and outgroup species.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{ML tree of *T. cirratus* and outgroup species.}
\end{figure}
higher than that within populations (27.98%). The results are in agreement with the findings of Fang [7], which confirmed that the differentiation mainly came from the population, but the proportion of genetic differentiation within the populations was still large. The genetic structure of the *T. cirratus* population has changed significantly after the invasion of CL, NL, and TL. We speculate that *T. cirratus* may adapt to the new invasive environment through the change of COI and Cytb gene.

Genetic diversity provides the potential for species to adapt to environmental change and is the key to long-term species development. Higher genetic diversity typically indicates that species are more adaptable to the environment, while low genetic diversity leads to lower fitness [33, 34]. Genomic analysis of the different populations of *T. cirratus* revealed that for the PR (*h* = 0.940, *π* = 0.033) and NR (*h* = 0.925, *π* = 0.034) populations, both haplotype diversity and nucleotide diversity were high; for the NL (*h* = 0.633, *π* = 0.001) and TL (*h* = 0.912, *π* = 0.006) populations, haplotype diversity was high but nucleotide diversity was low; and for the CL population (*h* = 0.000, *π* = 0.000), both haplotype diversity and nucleotide diversity were extremely low. The NR and PR are native populations, while the CL, TL, and NL are invasive populations, probably from the Yangtze River Estuary [7]. The genetic diversity of native *T. cirratus* populations was high, while that of the invasive populations was relatively low, owing to genetic drift during invasion. The extremely low genetic diversity of CL indicates that the *T. cirratus* population in the Chaohu Lake may come from only a few individuals.

Based on the molecular results of this study, it is hypothesized that *T. cirratus* experienced obvious genetic drift and low population diversity during invasion and that *T. cirratus* population outbreaks are less likely and might not have serious negative ecological impacts. The projection by Liang et al. [5] indicated that the suitable distribution area for *T. cirratus* will increase substantially under future climate change scenarios based only on the maximum entropy model. In the absence of molecular evidence, these results should be taken with caution.

### 4.2. Geographical Lineages

The two similar phylogenetic trees constructed by the BI and ML methods presented three major lineages. Lineage I contains a portion of haplotypes from NR; this could be an original and undisturbed population. Lineage II contains haplotypes from PR and NR that experienced a population bottleneck between 26 and 44 kya B.P.; this species may have been isolated due to sea level fluctuations caused by the Pleistocene glacial-interglacial period, and the current geographic pattern may be the result of secondary contact between the separated populations. Similarly, Liu et al. [35] identified the coexistence of three distinct taxa in the Northwest Pacific marginal seas of *Chelan haematocheilus* populations using mitochondrial DNA control region. Lineage III contains haplotypes from CL, NL, TL, and PR. According to the invasion history of *T. cirratus*, the separation time was not very long [2, 7]. The new mutations were fixed within population, but did not spread among populations. As a result, there is little variation in genetic distance within populations in different regions. In addition, the samples from the three lakes (non-native areas) are most likely from PR or locations with common genetic structure.

Fang [7] suggested dividing the *T. cirratus* geographical populations from Hainan Province (corresponds to the NR in present study), the Pearl River, and the Yangtze River basin into three separate valid species. The present study confirmed the existence of three phylogenetic lineages of *T. cirratus* in Chinese with large genetic differences. However, the haplotype network combined with phylogenetic trees showed that the populations in the Hainan Province, Pearl River, and Yangtze River basin did not form independent phylogenetic branches separately, so this study does not fully support Fang’s conclusion.

The neutral test, mismatch analysis, and BSP can infer demographic history based on molecular sequence samples [36]. According to the expansion time of lineage I and II population, the population expansion event was inferred to
have occurred about the Late Pleistocene, and the cause of the population expansion may be related to the climatic and geological events of the Late Pleistocene. As the Quaternary glacial period receded, the temperature rise led to the decrease of global ice volume and the rise of sea level, which brought many rich resources [37, 38] and provided suitable breeding conditions for *T. cirratus*, which promoted the rapid population growth and spatial expansion. Other populations of offshore fish in China have been similarly affected and show similar evidence in their DNA, such as the *Johnius grypotus* [39] and the *Periophthalmus novemradiatus* [40]. However, after 26 kya B.P., the effective population size of lineage III decreased, which may be related to the genetic drift that occurred in the lineage III population during the invasion. Individuals from the invasive populations may be unable to reveal the true population history of the Yangtze River [41].

4.3. Ancestral Haplotypes. Inferring ancestry from population genomic data is effective [42]. Hap1 is located at the centre of structure I, which is mainly composed of haplotypes from CL, NL, and TL. Hap25 is located at the centre of structure II, which is mainly composed of haplotypes from PR, and Hap58 is located at the centre of structure III, which is mainly composed of haplotypes from NR. Haplotypes located at the centre of each structure are considered ancestral haplotypes. Therefore, we conclude that Hap1, Hap25, and Hap58 are likely ancestral haplotypes.

4.4. Cryptic Species and Diversity. Molecular methods have improved our understanding of biodiversity in many taxa [43], including fishes [11, 44]. According to Wright's evaluation criteria [45], the genetic differentiation among intrabasin populations (CL, TL, and NL) was moderate ($F_{st} < 0.15$), while that among interbasin populations was high ($F_{st} > 0.15$). This suggests that partial geographic populations of *T. cirratus* may have cryptic species and diversity. After a long period of geographical isolation, gene exchange among species would be greatly reduced or even lost. With the passage of time, the intermediate transitional types generated at the beginning of the geographical isolation would gradually disappear, and the more divergent phylogenetic groups would emerge as subspecies or new species [46]. The haplotype diversity values in the PR and NR populations are quite high, approximately 0.94 and 0.93, respectively. Cryptic species may be hidden in the populations in the PR and NR. Lineage I and II may form subspecies or new species in the future. These results further support the idea of cryptic species and historical biogeography in Gobioidei [47]. Correspondingly, larger distribution ranges and sample sizes are required to unearth cryptic diversity in the *T. cirratus* populations in the future.

5. Conclusion

In the present study, we report molecular characterization and evolutionary analysis of *T. cirratus* in five geographical populations based on mitochondrial COI and Cytb gene sequences. The maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees cluster *T. cirratus* genes into three lineages. The $F_{st}$ and haplotype network analysis showed a high level of genetic differentiation among the populations of the Yangtze River basin, PR and NR, but they did not form independent phylogenetic branches separately. A change in effective population size of *T. cirratus* was investigated by Bayesian skyline plot. Hap1, Hap25, and Hap58 are likely the ancestral haplotypes. The partial geographic population of *T. cirratus* may have cryptic species and diversity. The study will be a vital road map for invasion biology pattern of *T. cirratus*. More samples covering larger distribution ranges are essential to illuminate the actual genetic diversity of this species in future studies. The results provide a further scientific basis for the study of fish taxonomy and fish diversity conservation as well as for the rational use of resources in fundamental research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Y.Y.L. designed the experiment and reviewed the manuscript. G.Q.Z. participated in manuscript writing. G.Q.Z. and C.C. were involved in molecular experiments and data analysis. W.X.L. and J.L. collected samples. Others contributed to molecular experimental sampling. All authors have reviewed and approved the final version of the manuscript.

Acknowledgments

This work was supported by the Natural Science Foundation of Anhui Province #1 under grant number 2008085QC105; Special Fund for Anhui Agriculture Research System #2 under grant Anhui Agricultural Science letter no. (2021)711, and Anhui Academy of Agricultural Sciences Wetland Ecology and Application Technology Innovation Team Project #3 under grant number 2021YL055.

References


