Identification of *Acrossocheilus fasciatus* (Teleostei: Cyprinidae) from Three Different Phylogeographic Sources using Morphology and COI DNA Barcoding

Haixia Zhang¹, Chang Hu¹, Wenli Duan¹, Luzhou Wang¹, Zhangjie Chu¹

**ABSTRACT**

**Background:** Accurate species delimitation is essential for phylogenetic, phylogeographic, ecological, conservation and biogeographic investigations. *Acrossocheilus* species classification has long been contentious, owing to their uncertain evolutionary relationships, high levels of morphological homoplasy and a lack of knowledge of the taxonomic validity of conventionally utilized morphological characters.

**Methods:** A final data set consisting of 102 mitochondrial cytochrome oxidase I (COI) gene sequences was analysed for K2P intraspecific and interspecific divergence in the current study, using reconstruction methods (maximum likelihood) and various models (ABGD, PTP, GMYC) to divide the sequences according to the pattern of genetic variation. We used landmark-based geometric morphometrics to identify specimens.

**Result:** Two taxa were identified according to the phylogenetic tree (maximum likelihood method) and three different models (ABGD, PTP, GMYC). The molecular taxonomy of *Acrossocheilus* agreed with the morphological identification of specimens from the same collection locality. Therefore, the COI gene and geometric morphometrics were effective in species identification of *A. fasciatus*.

**Key words:** *Acrossocheilus fasciatus*, Biodiversity, COI, Morphological taxonomy, Species identification.

**INTRODUCTION**

The genus *Acrossocheilus* (Oshima, 1919), belonging to Cypriniformes, Cyprinidae and Barbinae, is mainly distributed in the Yangtze River and its southern water system, including Taiwan Province and Hainan Island, with a few species found in Vietnam and Laos. It is a group of small and medium-sized freshwater fishes living in rapids (Hou et al., 2020). There are 47 species and subspecies in the genus *Acrossocheilus* according to the FishBase database (https://www.fishbase.cn/search.php). Excluding synonyms, there are 26 genuine species, with 20 species and subspecies found in China. This large number of species makes it possible for populations of the same species to differ morphologically owing to variations in the environment, making it exceedingly challenging to identify a particular species. Additionally, there is currently no commonly accepted system of morphological categorization (Laakmann et al., 2017). The significant degree of morphological trait overlap across species, in general, further amplifies this complexity, as identification typically relies on a small number of inconsistent features.

Reliable information on species richness is vital for any study of biodiversity and conservation strategy, although it is sometimes impossible to differentiate one species from another based on physical characteristics that are quite similar. In this context, accurate species identification is the first and most crucial stage in the application of conservation strategies and the sustainable exploitation of natural resources, especially in light of the present accelerating biodiversity crisis resulting from human activities.

DNA barcoding is a molecular biology method for species identification based on DNA sequences (Palasio et al., 2017). It is frequently used in species identification and genetic diversity research since it does not rely on the integrity of the stored samples, is unaffected by the developmental stage of the organism and has a better level of accuracy when identifying comparable species. The mitochondrial cytochrome c oxidase I subunit (COI) gene is regarded as the perfect DNA barcode since it has a sufficient evolutionary rate and is simple to amplify. Studies have demonstrated that the COI gene has a high level of accuracy for identifying fish species and can discriminate between approximately 98% of marine fishes and 93% of freshwater fishes.

Using the COI gene, we performed a molecular phylogenetic analysis to investigate the genetic variations and evolutionary linkages between the species according
to molecular, morphological and phylogenetic analyses. The results can serve as a reference for taxonomic research on the genus *Acrossocheilus*, for more in-depth conservation management and for the sustainable use of fish resources (Zheng et al., 2018).

**MATERIALS AND METHODS**

**Sample collection**

We collected 120 samples of *A. fasciatus* from three municipalities in the provinces of Anhui and Zhejiang in China (Fig 1). These specimens were collected between June 2021 and December 2021. The experiment was carried out in the laboratory of Zhejiang Ocean University in November and December 2021. Muscles were extracted from the live animals and kept in 95% ethanol for subsequent genetic research (Tsoupas et al., 2022).

**Morphological studies**

The samples analysed included a total of 120 individuals from three different geographical locations and shortly after obtaining the samples, they were euthanized, photographed and sexed by gonadal examination. The scales (Microscope Olympus BX43 Olympus Life Sciences (Tokyo) were used to establish the age of the individuals and all samples were roughly one year old. Using the TpsDig20 program, each fish was photographed from the side and the landmarks in each image were digitally enhanced (Fig 2) (Rohlf 2015). Generalized Procrustes analysis (GPA) was used to align and superimpose landmark configurations to cancel out the impacts of different body sizes, positions and orientations (Fig 3). The overlaid residuals were then evaluated using the thin-plate spline (TPS) interpolation function.

The MorphoJ program (Klingenberg, 2011) was utilized based on the covariance matrix to minimize the dimensionality of the data for each participant prior to performing a comparison analysis (Zelditch et al., 2004). This made it possible to determine if these impacts on the sample scores along the two major principal components (PCs) were significant.

**DNA extraction, amplification and sequencing**

The DNeasy Tissue Kit was used to isolate DNA (Qiagen). Regarding primers for PCR, it was necessary to obtain several COI gene fragments from fish samples; therefore, we used the universal primers COIF: 5'-TGTAAAACGACGGCCAGTCCTGTGGAATYACDCGCTGAT and COIR: 5'-CAGGAAACAGCTATGACNYTCNGGRTGNCCRAAGAA (Huang et al., 2015). The following reaction conditions were used for polymerase chain reaction (PCR): 10-100 ng of DNA, 2 µl of each dNTP, 1.6 µl of each primer and 0.3 µl of Taq DNA polymerase in the given reaction buffer. There were 35 cycles of 0.5 min at 94°C, 0.5 min at 55°C and 0.75 min at 72°C and a final extension step of 7 min at 72°C (Hoddle et al., 2008). The first step of denaturation lasted for 4 min at 94°C. The Qiagen purification kit was used to purify the PCR products before being sequenced on the ABI 3100 automated sequencer (Applied Biosystems) at the Biotechnology Research Institute Biotechnology Center.

**Molecular data analysis**

After being adjusted with CHROMAS (Technelysium Pty Ltd.), the Schentropic profiles acquired from the sequencing of each sample were compared to the CLUSTALX version. The BIOEDIT version was used to modify the aligned sequences (Hall 1999). The Kimura 2-parameter (K2P) technique and MEGA7 (Molecular Evolutionary Genetics Analysis) software (Tamura et al., 2013) were used to compute the intra- and interspecific genetic distances and maximum likelihood analysis was also carried out using MEGA7. The HKY model of sequence evolution was chosen as the best model for further studies in the maximum likelihood (ML) analysis using the Akaike information criteria implemented in MODELTEST 23. Utilizing ABGD (Puillandre et al., 2012), PTP (Zhang et al., 2013) and GMYC methodologies (Rambaut et al., 2018), a barcoding gap study was carried out. These methodologies are available at the following URL: http://www.abisnvjussieufr/public/.abgd/. The server utilizes the default settings separately for http://species.hits.org and http://speciesh-itsorg/gmyc/web.

Fig 1: Sampling sites in this analysis. (QS: Qianshan; HS: Huangshan; YH: Yunhe).
RESULTS AND DISCUSSION

The PCA biplot summarizes the overall variety in body shape. The first PC (PC1) explained 31.2% of the variation and the second PC (PC2) explained 18.7% of the variation, for a total of 49.9% of the variation described by these two components. The distribution of PC1 in the scatter plot with PC2 shows that the Huangshan and Qianshan groups basically overlap in one range, with an intersection point with the Yunhe group (Fig 4).

Molecular analysis

Since the morphological characters were not effective in differentiating A. fasciatus, the results of species delimitation based on three different molecular delimitation methods (GMYC, PTP and ABGD) were used to delimit A. fasciatus based on the COI gene. The results showed that the molecular delimitation analysis yielded generally consistent results, except for the ABGD analysis with the interspecific genetic distance (P) set to 0.001 and 0.0599, which delimited 8 MOTUs and 1 MOTU (Fig 5). The remaining molecular delimitation analyses yielded roughly the same results, with 2 MOTUs. According to the interspecific genetic distance (P), the ABGD analysis divided species of A. fasciatus into various numbers of MOTUs. If the interspecific genetic distance (P) is too high, the entire data set is treated as one species, indicating excessive division. In contrast, if the interspecific genetic distance (P) is too low, only the same sequences will be treated as belonging to the same species. Puillandre advised adopting an interspecific genetic distance (P) of 0.01 to delimit species in this situation because findings from ABGD analysis will be as consistent as feasible with those from other approaches when the interspecific genetic distance (P) is 0.01. In our study, ABGD analysis delineated a total of two MOTUs when the interspecific genetic distance (P) was 0.01, which was consistent with the results of our PTP analysis; therefore, we decided to use the delineation result under an interspecific genetic distance (P) of 0.01 as the final result of ABGD analysis in this study. In GMYC analysis, either single or multiple threshold approaches may result in the over-demarcation of species, but in our analysis, multi threshold GMYC analysis yielded relatively reliable demarcation results, as they were largely consistent with the results of ABGD and PTP analyses. Ultimately, considering the results of all the different molecular delimitation analyses, a total of 2 MOTUs were delineated, with HS and QS treated as one MOTU and YH as a separate MOTU.

Finally 102 effective sequences were obtained by amplification and sequencing, with HS and QS representing the named species A. fasciatus. Amplification of the COI gene yielded a normalized 677 bp fragment and nucleotide composition analysis revealed an average nucleotide composition of 25.56% adenine (A), 27.62% thymine (T), 18.17% guanine (G) and 28.65% cytosine (C).

Two significant branches were found by maximum likelihood (ML) using COI-based sequencing (with the exception of outgroup) (Fig 6).
Identification of *Acrossocheilus fasciatus* (Teleostei: Cyprinidae) from Three Different Phylogeographic Sources using Morphology

**ESU delimitation**

In this study, we used molecular and morphological methods for the phylogeographic analysis of *A. fasciatus* from three localities. Molecular analyses showed consistent results with the ML tree and three species identification methods based on different principles, indicating two evolutionarily significant units (ESUs), in terms of both barcode spacing and differential analysis of phylogenetic processes. HS and QS are more closely related to each other according to intraspecific genetic distances and can be considered to belong to the same species. The morphological characteristics determined in this study are consistent with the results of molecular analysis and the morphological analysis revealed that HS and QS were more closely related in terms of PCs.

In determining the origin of fish based on morphology (Fleming *et al*., 1994), the accuracy can reach 100%. In the current study, morphometric analysis based on geometric morphometry was applied to populations. Geometric morphometric indicators can be a useful tool for measuring fish community diversity (Chandran *et al*., 2022). This is consistent with the findings of actual fish and gastropod assemblage structures (McClain *et al*., 2004; Lombarte *et al*., 2012). In the present study, differences in overall body morphology were present in *A. fasciatus* and there was significant differentiation between the fish of the two taxonomic units in the PCA. The use of morphological methods alone is more challenging for the accurate identification of fish species, highlighting the urgent need to use molecular methods in taxonomic investigations (Rathnasuriya *et al*., 2021).

The use of the COI gene as a tag in DNA barcoding to identify species, particularly fish species, has recently gained increasing attention (McCusker *et al*., 2013; Knebelsberger *et al*., 2014). The common pattern of variation observed in numerous species is one of the key grounds for choosing COI as the standard treatment barcode gene, with both marked divergence and a lack of overlap between intraspecific and interspecific genetic distances (Hebert *et al*., 2003).

Of the 3 geographic locations investigated in this study, only 1 region had species with a K2P value exceeding 2%. Sequence diversity among species from two regions is less than 2%, indicating no increase in genetic variety in

---

**Fig 4:** A principal component biplot displaying the variation in overall body shape for the HS, QS and YH groups.

**Fig 5:** The number of MOTUs estimated from automatic barcode gap discovery.
Identification of *Acrossocheilus fasciatus* (Teleostei: Cyprinidae) from Three Different Phylogeographic Sources using Morphology...

**Systematic geography**

Geographically, HS and QS are separated by the Yangtze River. Some areas in the Yangtze River’s middle and lower reaches are blocked, while QS is located in a section of the river that is not blocked (You et al., 2017). A study conducted by Shang et al. (2022) showed that the Yangtze River did not block their gene exchange. Surface currents may carry these species from hatching sites to nursery areas with suitable feeding conditions (Yang et al., 2018). All fish populations that live in places with stable existing systems have a genetic predisposition to survive within geographic regions (Ai et al., 2021). The authors described the life cycles of upstream hatching areas and downstream nursery grounds and the theory’s general applicability has been demonstrated to apply to numerous fish populations in multiple geographical regions, including the North Atlantic as well as the South African coast adjoining the South Atlantic and Indian Oceans (Hutchings et al., 2002; Kantharajan et al., 2022). Hutchings et al. (2002) discovered that the majority of pelagic and demersal fishes have evolved highly selective reproduction methods, which include fish migrating to spawn in upstream areas, where the offspring are subsequently transported by ocean currents to suitable nursery locations. HS and YH are located on both sides of Huanghai Mountain, a famous scenic area and a national tourist attraction. The results of molecular and morphological analyses were consistent, with YH on a separate branch. Thus, Huanghai Mountain, as a watershed, blocked the genetic exchange of YH with HS and QS.

**Conservation implications**

Our results emphasize the significance of species delimitation using an integrated framework that might improve Cyprinidae taxonomy and classification, which is critical for freshwater fish phylogeny, phylogeography, ecology, conservation and biogeography research. We believe that by using exploratory delimitation analysis, which treats a probable species as a guide and contrasts it on the basis of morphology and other methodologies, it will be possible to overcome challenges in identifying boundaries at the species or genus level in Cyprinidae taxonomy.

**Conclusion**

The external morphological characteristics of species have always played an important role in the classification of species. At first, it was widely used in the classification of species. However, in the process of evolution, the morphology of species is easily influenced by the environment and tends to evolve. For cryptic species with very similar morphology, the purpose of identification and
Identification cannot be achieved simply by using morphological methods and the data of molecular phylogeny and other research needs to be combined. The classification of the Acrossocheilus is chaotic and the morphology is extremely chaotic. If the traditional morphology is used for classification, different scholars may have different results due to different experiences. In this study, the DNA barcode technology of mitochondrial cytochrome oxidase I (COI) gene sequence was used and the tree reconstruction method (maximum likelihood method ML) and different species definition models (ABGD, GMYC and PTP) were used to divide the sequences according to the genetic variation model and the intra-species and inter-species differences of K2P were analyzed. Geometrical measurement standards were used to distinguish samples and the morphological identification of samples collected in the same place was consistent with the molecular taxonomy study of A. fasciatus. The results showed that no cryptic species had been found yet.

Conflict of interest: None.

REFERENCES


