

## Research Article

# Species Identification of Marine Fishes in China with DNA Barcoding

**Junbin Zhang**

*College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China*

Correspondence should be addressed to Junbin Zhang, jbzhang30@163.com

Received 10 December 2010; Accepted 27 February 2011

Copyright © 2011 Junbin Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

DNA barcoding is a molecular method that uses a short standardized DNA sequence as a species identification tool. In this study, the standard 652 base-pair region of the mitochondrial cytochrome oxidase subunit I gene (COI) was sequenced in marine fish specimens captured in China. The average genetic distance was 50-fold higher between species than within species, as Kimura two parameter (K2P) genetic distances averaged 15.742% among congeners and only 0.319% for intraspecific individuals. There are no overlaps of pairwise genetic variations between conspecific and interspecific comparisons apart from the genera *Pampus* in which the introgressive hybridization was detected. High efficiency of species identification was demonstrated in the present study by DNA barcoding. Due to the incidence of cryptic species, an assumed threshold is suggested to expedite discovering of new species and biodiversity, especially involving biotas of few studies.

## 1. Introduction

Fishes are important animal protein sources for human beings, and they are frequently used in complementary and alternative medicine/traditional medicine (CAM/TM). The delimitation and recognition of fish species is not only of interest for taxonomy and systematics, but also a requirement in management of fisheries, authentication of food products, and identification of CAM/TM materials [1–3].

Due to the complexity and limitations of morphological characters used in traditional taxonomy, several PCR-based methods of genotype analysis have been developed for the identification of fish species, particularly for eggs, larvae, and commercial products. Sequence analysis of species-specific DNA fragments (often mitochondrial or ribosomal genes) and multiplex PCR of species-conserved DNA fragments are efficient for fish species identification [4–10]. However, these molecular methods are limited to particular known species and are not easily applicable to a wide range of taxa. Therefore, Hebert et al. advocated using a standard DNA sequence that is DNA barcoding to identify species and uncover biological diversity [11, 12]. For many animal taxa, sequence divergences within the 5' region of the mitochondrial cytochrome oxidase subunit I (COI) gene were much greater between species than within them, and this in turn suggests that the approach is widely applicable

across phylogenetically distant animal groups [12, 13]. To date, some published papers explicitly address that COI barcodes effectively discriminate different species for a variety of organisms [14–23]. However, several scientists express concerns that species identification based on variations of single mitochondrial gene fragment may remain incorrect or ambiguous assignments, particularly in cases of possible mitochondrial polyphyly or paraphyly [24, 25]. In the current study, we test the efficacy of DNA barcoding in marine fishes of China. The sea area of China is part of the Indo-West Pacific Ocean, which is regarded as the center of the world's marine biodiversity [26]. Highly species-rich biotas are particularly attractive to test the reliability and efficiency of DNA barcoding.

## 2. Material and Methods

The majority of fish specimens were captured with the draw net at 20 localities along the coast of China (collection information available at <http://www.barcodinglife.org/>). A total of 329 specimens from one hundred species of fish were collected. Vouchers were deposited in the South China Sea Institute of Oceanography, Chinese Academy of Sciences, and all specimens were preserved in 70% ethanol. Tissue samples were dissected from the dorsal muscle, and genomic DNA was extracted according to the standard Barcode of Life

protocol [27]. Firstly, fragments of the 5' region of the mitochondrial COI gene were PCR-amplified using C.FishF1t1/ C.FishR1t1 primer cocktails [28]. The cocktail C.FishF1t1 contained two primers (FishF2.t1/VF2.t1), and C.FishR1t1 also contained two primers (FishR2.t1/ FR1d.t1). All PCR primers were tailed with M13 sequences to facilitate sequencing of products. The nucleotide sequences of the primers were

FishF2.t1: \*5'-TGTAAAACGACGGCCAGTTCGACT-AATCATAAAGATATCGGCAC-3'.

VF2.t1: \*5'-TGTAAAACGACGGCCAGTCAACCA-ACCACAAAGACATTGGCAC-3'.

FishR2.t1: \*\*5'-CAGGAAACAGCTATGACACTTC-AGGGTGACCGAAGAATCAGAA-3'.

FR1d.t1: \*\*5'-CAGGAAACAGCTATGACACCTCA-GGGTGTCCGAARAAYCARAA-3'.

\*The M13F primer sequence is underlined; \*\*the M13R primer sequence is underlined.

PCR reactions were carried out in 96-well plates using Mastercycler Eppendorf gradient thermal cyclers (Brinkmann Instruments, Inc.). The reaction mixture of 825  $\mu$ l water, 125  $\mu$ l 10 $\times$  buffer, 62.5  $\mu$ l MgCl<sub>2</sub> (25 mM), 6.25  $\mu$ l dNTP (10 mM), 6.25  $\mu$ l each primer (0.01 mM), and 6.25  $\mu$ l Taq DNA polymerase (5 U/ $\mu$ l) was prepared for 96 wells of each plate, in which each well contained 10.5  $\mu$ l mixture and 2  $\mu$ l genomic DNA. Thermocycling comprised an initial step of 2 min at 95°C and 35 cycles of 30 sec at 94°C, 40 sec at 52°C, and 1 min at 72°C, with a final extension at 72°C for 10 min. Amplicons were visualized on 2% agarose E-Gel 96-well system (Invitrogen). PCR products were amplified again with the primers M13F (5'-TGTAAAACGACGGCCAGT-3') and M13R (5'-CAGGAAACAGCTATGAC-3'), respectively, using the BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.). Thermocycling conditions were as follows: an initial step of 2 min at 96°C and 35 cycles of 30 sec at 96°C, 15 sec at 55°C, and 4 min at 60°C. Sequencing was performed on an ABI 3730 capillary sequencer according to manufacturer's instructions.

For specimens that failed to yield sequences using the primer combinations above, a second round of PCR using the alternative C.VF1LFt1/ C.VR1LRt1 primer combination was carried out. C.VF1LFt1 consisted of four primers (VF1.t1/VF1d.t1/LepF1.t1/VFli.t1), and C.VR1LRt1 also comprised four primers (VR1.t1/VR1d.t1/LepR1.t1/VRli.t1) [28].

VF1.t1: \*5'-TGTAAAACGACGGCCAGTTCCTCAA-CCAACCACAAAGACATTGG-3'.

VF1d.t1: \*5'-TGTAAAACGACGGCCAGTTCCTCAA-CCAACCACAARGAYATYGG-3'.

LepF1.t1: \*5'-TGTAAAACGACGGCCAGTATTCACCAATCATAAAGATATTGG-3'.

VFli.t1: \*5'-TGTAAAACGACGGCCAGTTCCTCAA-CCAACCAIAAIGAIATIGG-3'.

VR1.t1: \*\*5'-CAGGAAACAGCTATGACTAGACT-TCTGGGTGGCCRAARAAYCA-3'.

VR1d.t1: \*\*5'-CAGGAAACAGCTATGACTAGACT-TCTGGGTGGCCAAAGAATCA-3'.

LepR1.t1: \*\*5'-CAGGAAACAGCTATGACTAAAC-TTCTGGATGTCCAAAAAATCA-3'.

VRli.t1: \*\*5'-CAGGAAACAGCTATGACTAGACT-TCTGGGTGICCIAAIAAICA-3'.

\*The M13F primer sequence is underlined; \*\*the M13R primer sequence is underlined.

The thermocycling protocol used was 1 min at 95°C and 35 cycles of 30 sec at 94°C, 40 sec at 50°C, and 1 min at 72°C, with a final extension at 72°C for 10 min. Sequencing PCR and sequencing followed above procedure.

DNA sequences were aligned with SEQSCAPE v.2.5 software (Applied Biosystems, Inc.). Sequence divergences were calculated using the Kimura two parameter (K2P) distance model [29], and unrooted NJ trees based on K2P distances were created in MEGA software [30]. In the chosen taxonomic group, phylogenetic analysis was carried out in PAUP 4.010b using the maximum parsimony (MP) method, with 1,000 replications of the full heuristic search.

The following categories of K2P distances were calculated: intraspecific distances (S), interspecies within the congener (G), and interspecies from different genus but within intrafamily (F). These values were plotted using the boxplot representation of R. Boxplots [31] in SPSS 11.5 software (SPSS Inc., Chicago, IL, USA). Only for families containing 2 or more genera, separate boxplot was constructed for the sake of comparisons among taxonomic categories. Boxplots describe median (central bar), interquartile range (IQR: between upper (Q3) and low (Q1) quartile), values lying within 1.5 $\times$  IQR beneath Q1 or 1.5 $\times$  above Q3 ("whiskers"), and extreme values (outliers). Mann-Whitney tests were performed between S, G, and F distributions to estimate the overlap among taxonomic ranks.

### 3. Results

A total of 329 specimens were analyzed, from which 321 sequences (all >500 bp) belonging to 121 species (another species was identified to the genus level) were ultimately obtained (GenBank accession numbers: EF607296-EF607616). These species cover the majority of fishes living in the coastline of the South China Sea. All sequences were aligned with a consensus length of 652 bp, and no insertions, deletions, or stop codons were observed in any sequence. However, multiple haplotypes were detected for some species.

Except for *Acentrogobius caninus*, *Scomber japonicus*, *Terapon jarbua*, *Upeneus sulphureus*, *Elops hawaiiensis*, *Gymnothorax pseudothyroideus*, *Dendrophysa russelii*, and *Pennahia anea* (which reached the maximum value of 2.02%), intraspecific genetic distances were generally below 1%, and some decreased to zero (between some intraspecific individuals of *Thryssa setirostris*, *Paraperis ommatura*, *Scatophagus argus*, etc.).

The mean intraspecies K2P (Kimura two-parameter) distance was 0.319%; the distance increased sharply to 15.742% among individuals of congeneric species. Overall,

TABLE 1: Genetic divergences (percentage, K2P distance) within various taxonomic levels. Data are based on 321 sequences (>500 bp) from 122 species.

Comparisons within	Taxa	Number of comparisons	Mean	Median	Minimum	Maximum	s.e. <sup>#</sup>
Species	121	453	0.319	0.150	0	2.021*	0.018
Genus	85	397	15.742	16.490	0.154**	25.189	0.292
Family	55	848	20.199	19.850	11.532	34.333	0.134
Order	15	17881	24.656	—	12.923	39.627	0.024
Class	2	29262	25.225	—	15.730	40.800	0.016

\* Pennahia anea; \*\* Pampus argenteus versus Pampus cinereus.

<sup>#</sup>Standard error.

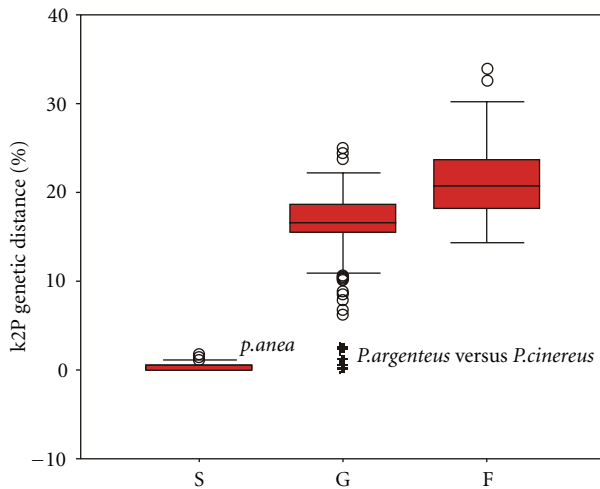


FIGURE 1: Box plots of K2P distances. IQR: interval into which the “central” 50% of the data fall. Black bar in the box indicates the median. Circle: “mild outlier” and asterisks: “extreme outliers”. Extreme outliers are discussed in the text.

the average genetic distance among congeneric species is nearly 50-fold higher than that among individuals within species. For the higher taxonomic ranks (family, order, and class), mean pairwise genetic distances increased gradually and reached 20.199%, 24.656%, and 25.225%, respectively (Table 1). Standard errors for K2P genetic distances were small, and values of the mean and median were close within different taxonomic ranks (Table 1). This indicates fluctuations of K2P genetic distances tend to be convergent (Figures 1 and 2).

In the unrooted NJ (neighbour-joining) tree (Figure 3), three specimens of *Pampus argenteus* were grouped together and contained within the cluster of *Pampus cinereus*. These *Pampus argenteus* specimens were collected in the same site off the west coast of the South China Sea, and were difficult to identify because of their complex morphological characteristics (available at <http://www.barcodinglife.org/>). They possessed combined characteristics of *Pampus cinereus* and *Pampus argenteus*: the asymmetrical tail of *Pampus cinereus* and silver color of *Pampus argenteus*. If the suspicious congeneric K2P distances in the genera *Pampus* are excluded (the extreme outliers in Figure 1), the pairwise genetic

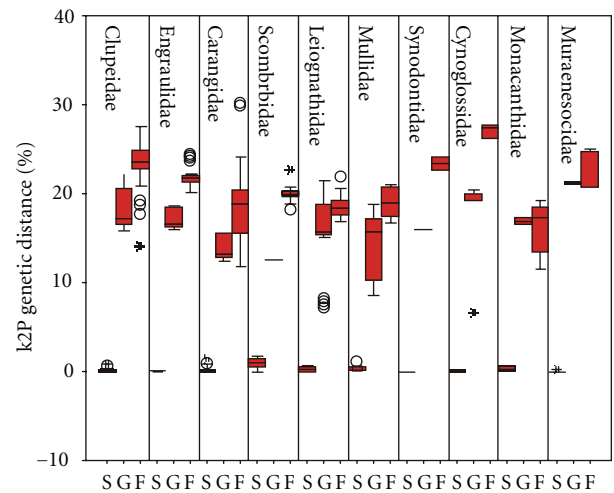


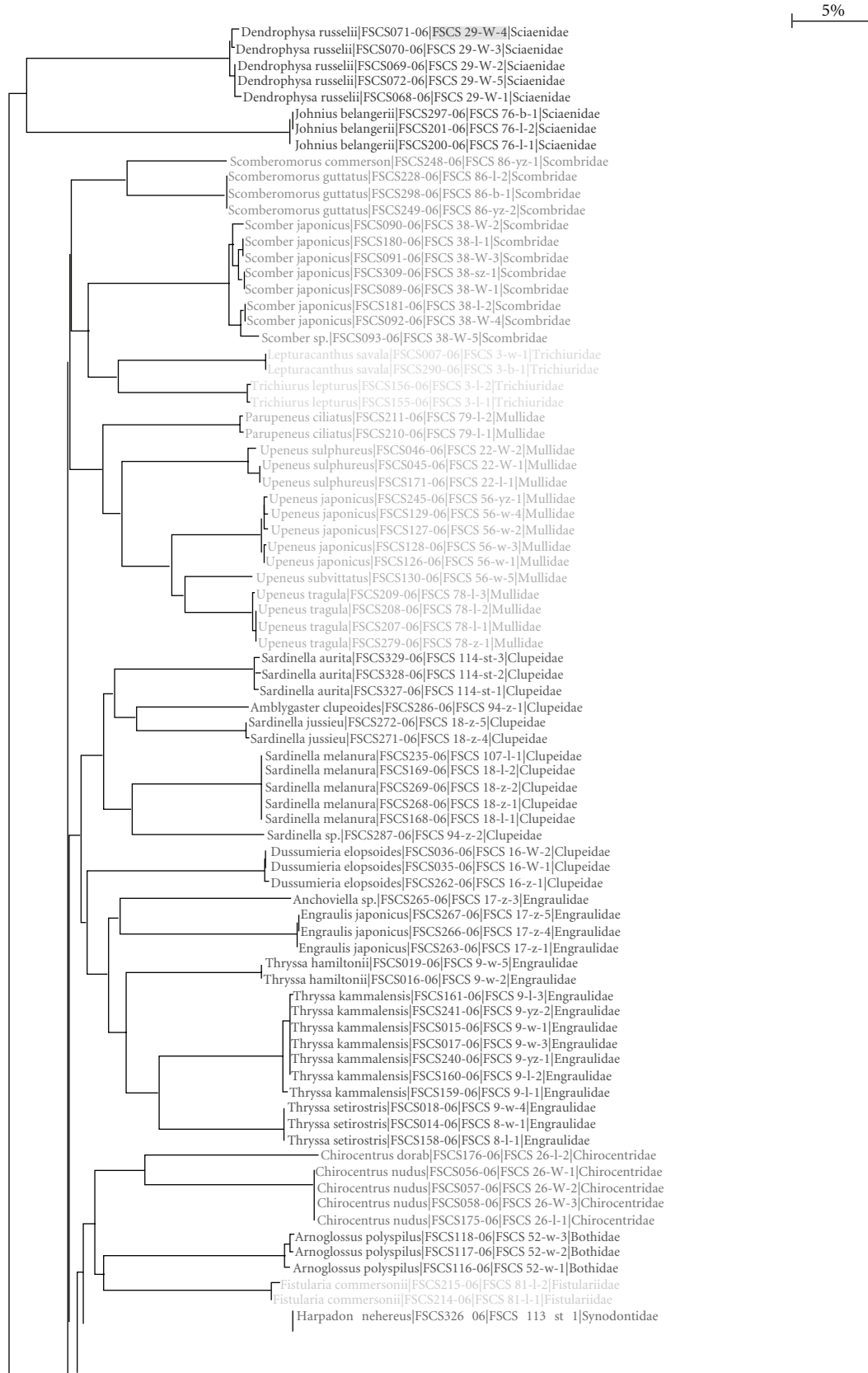
FIGURE 2: Boxplot distributions of S, G, and F. Intra-species (S), interspecies among congeneric species (G), and intergenera but intrafamily (F) K2P distances for different families.

divergences among congeneric species are above 10%. There are no overlaps between intraspecific and congeneric K2P distances within the same family (Figure 3).

At the species level, all COI sequences clustered in monophyletic species units. At the family level, there were paraphyletic clusters for three families (Carangidae, Gobiidae, and Ariidae) (Figure 3), though over 98% of specimens fell into the expected division of families. Intrafamily K2P distances (F) were generally higher than congeneric (G) distances, which were definitely higher than intraspecific (S) distances (Table 1, all Mann-Whitney tests were highly significant, P value <10<sup>-6</sup>). However, overlaps between F and G distances were observed in Clupeidae, Carangidae, Mullidae, and Muraenesocidae.

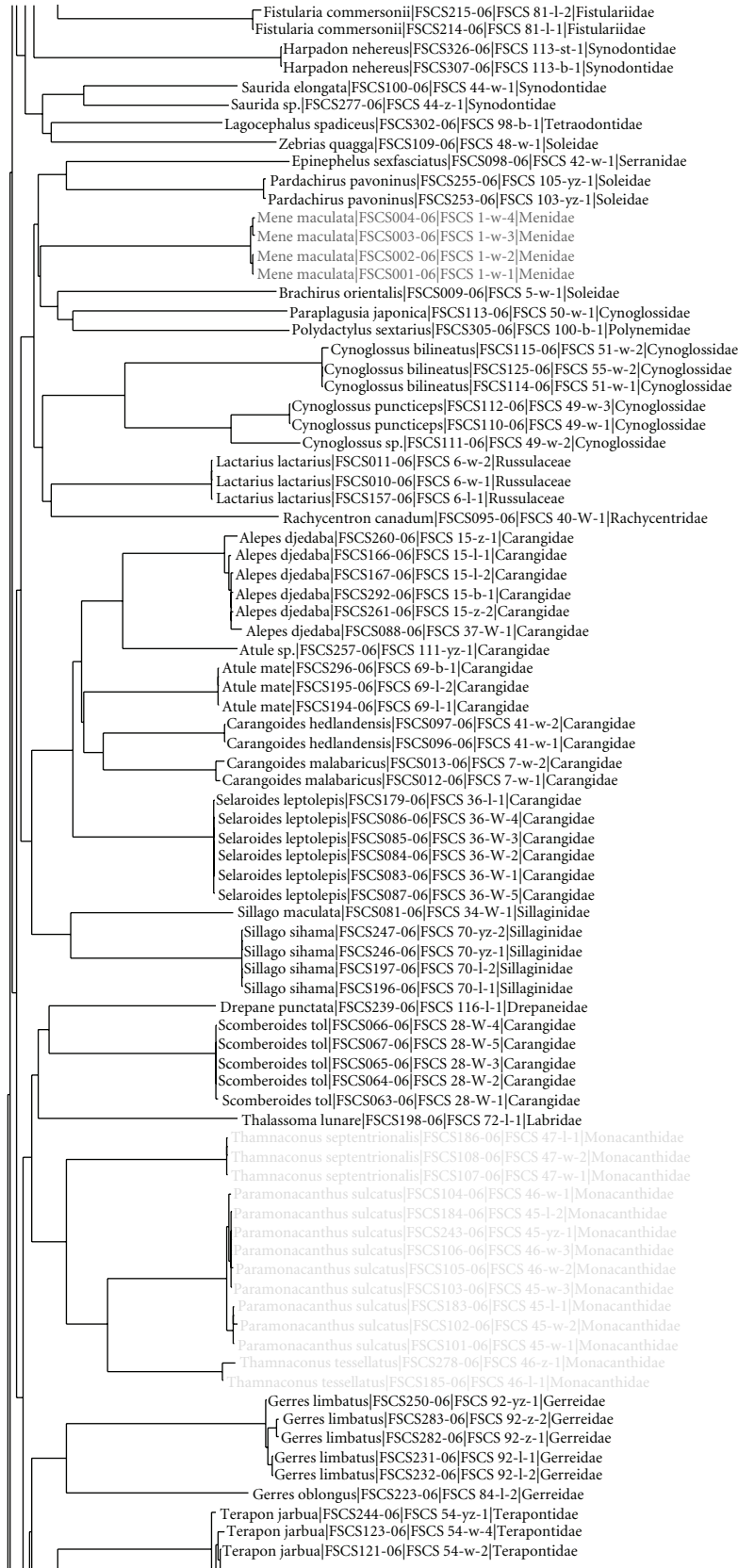
#### 4. Discussion

In morphological taxonomy, characters are delimited usually without any explicit criteria for character selection or coding, and morphological data sets have the potential to be quite arbitrary. For example, morphologists do not generally report their criteria for including or excluding characters,



(a)

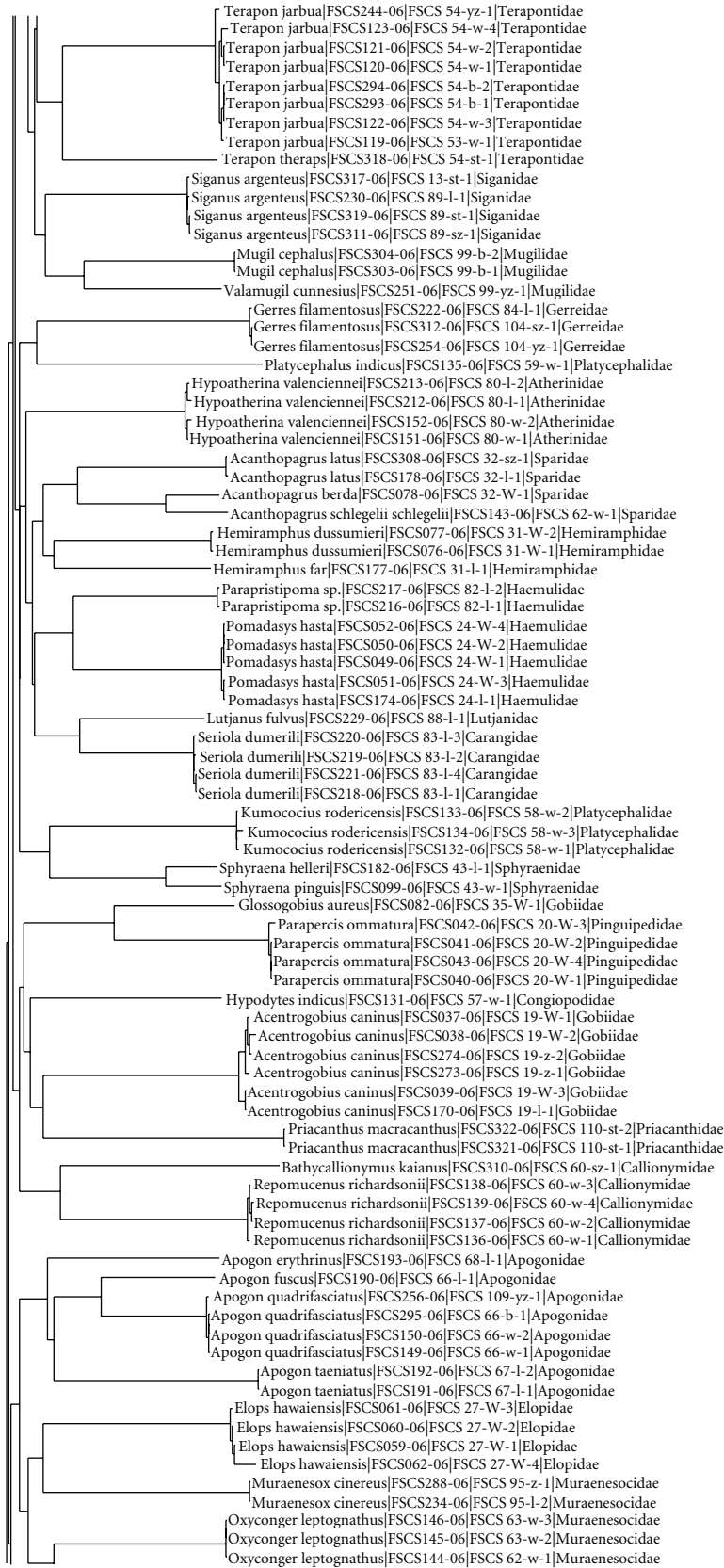
FIGURE 3: Continued.



(b)

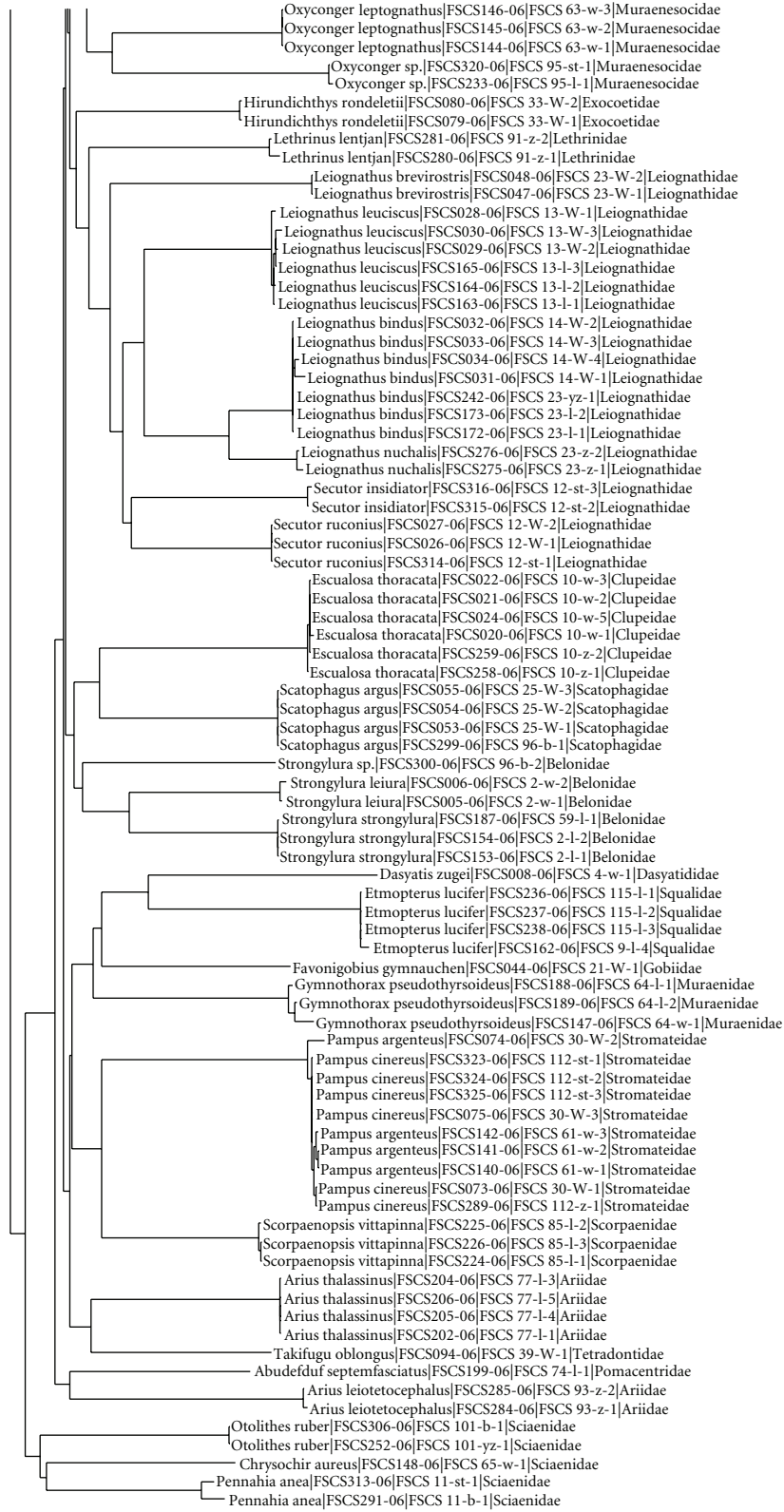
FIGURE 3: Continued.





(c)

FIGURE 3: Continued.



(d)

FIGURE 3: Neighbor-joining (NJ) tree of COI sequences. Scale: 5% K2P distance. The first numbers following species names are the process IDs, and the latter are the sample IDs.

and when criteria are given, they vary considerably among studies [32]. Thus, it is not surprising that there are so many synonyms for organisms [33], and an objective, rigorous species delimitation according to explicit criteria is therefore necessary for many taxonomic studies [34]. While DNA barcoding provides taxonomic identification for a specimen, the accuracy of such an assignment depends on whether species are monophyletic with respect to sequence variations of the COI gene. That is, individuals of a given species are more closely related to all other conspecifics than to any member of other species. Except for the hybridized specimens in the genus *Pampus*, there are no overlaps between genetic variations of S and G (Figure 1).

The factors responsible for deviations from taxonomic monophyly may be varied and complex [35]; one potential cause of species-level polyphyly is the occasional mating between distinct species, resulting in hybrid offspring carrying a mixture of genes from both parent species. Furthermore, mitochondrial genes are generally subjected to introgression more frequently than nuclear ones, and introgression also leads to phylogenetic paraphyly [35–38], like the hybridization between *Pampus argentatus* and *Pampus cinereus* in this study. In such cases, combinations of morphological and genotypic data are needed for species assignment of hybrids.

Biological mechanisms, water dynamics, or historical events may cause deep genetic structuring of populations in marine species [26, 39]. Many explanations for genetic population structuring on local and regional scales involve behaviors such as the adoption of pelagic early life stages and movement over broad geographic ranges, and these factors are theoretically associated with gene flow [40–42]. For many marine fishes, there is a lack of phylogeographic structure among populations [43, 44]; in this study, for individuals from long distance localities, some intraspecific genetic variations reduced to zero within families Carangidae, Sciaenidae, and Mullidae. However, some pairwise K2P distances exceeded 1.00% within the coastal species such as *Acentrogobius caninus*, *Scomber japonicus*, *Terapon jarbua*, *Upeneus sulphureus*, *Elops hawaiiensis*, *Gymnothorax pseudothyrsoides*, and *Dendrophysa russelii*. It implied that biological mechanisms were responsible for the fluctuation of intraspecific genetic divergences in marine fishes.

The neighbor-joining method was originally employed in this study for species identification, but some phylogenetic information was also revealed by the dendrogram, and over 98% of specimens were allocated into different families without polyphyly/paraphyly in the NJ tree (Figure 3). However, DNA barcoding is independent of the way the taxonomy has been built, and it cannot be regarded as the “taxonomic” tag [45]. DNA barcoding is no substitute for taxonomy Ebach and Holdrege [46], and a great deal of work is needed to bring about the reconciliation between traditional and molecular taxonomy. It is unfeasible to build the phylogeny of fishes only based on mitochondrial DNA fragments. Polyphyly/paraphyly in the NJ tree probably results from “bad taxonomy” when named species fail to identify the genetic limits of separate evolutionary entities, particularly for perplexing taxa involving cryptic species [47]. If we

cannot set a threshold of the genetic variation in species delimitation, we find ourselves sunk in the dilemma facing new or cryptic species. On the one hand, the morphological taxonomy cannot give a definite identification. On the other hand, we cannot claim that it may be a new species based on molecular analysis without the species delimitation. An assumed threshold is helpful to expedite discovery of new species and biodiversity, especially in dealing with little-studied biotas, although a single, uniform threshold for species delimitation seems arbitrary because rates of molecular evolution vary widely within and among lineages [24, 25, 48].

## Acknowledgments

Thanks are due to Chinese National Funding U0633007 and 40776089 and CAS Founding KZCX2-YW-213.

## References

- [1] R. S. Rasmussen, M. T. Morrissey, and P. D. N. Hebert, “DNA barcoding of commercially important salmon and trout species (*oncorhynchus* and *salmo*) from North America,” *Journal of Agricultural and Food Chemistry*, vol. 57, no. 18, pp. 8379–8385, 2009.
- [2] B. C. Victor, R. Hanner, M. Shivji, J. Hyde, and C. Caldwell, “Identification of the larval and juvenile stages of the Cubera Snapper, *Lutjanus cyanoptemus*, using DNA barcoding,” *Zootaxa*, no. 2215, pp. 24–36, 2009.
- [3] B. Patwardhan, D. Warude, P. Pushpangadan, and N. Bhatt, “Ayurveda and traditional Chinese medicine: a comparative overview,” *Evidence-Based Complementary and Alternative Medicine*, vol. 2, no. 4, pp. 465–473, 2005.
- [4] S. E. Bartlett and W. S. Davidson, “Identification of *Thunnus tuna* species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome b genes,” *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 48, pp. 309–317, 1991.
- [5] A. Rocha-Olivares, “Multiplex haplotype-specific PCR: a new approach for species identification of the early life stages of rockfishes of the species-rich genus *Sebastes* Cuvier,” *Journal of Experimental Marine Biology and Ecology*, vol. 231, no. 2, pp. 279–290, 1998.
- [6] A. J. Gharrett, A. K. Gray, and J. Heifetz, “Identification of rockfish (*Sebastes* spp.) by restriction site analysis of the mitochondrial ND-3/ND-4 and 12S/16S rRNA gene regions,” *Fishery Bulletin*, vol. 99, no. 1, pp. 49–62, 2001.
- [7] B. Horstkotte and H. Rehbein, “Fish species identification by means of restriction fragment length polymorphism and high-performance liquid chromatography,” *Journal of Food Science*, vol. 68, no. 9, pp. 2658–2666, 2003.
- [8] C. J. Noell, S. Donnellan, R. Foster, and L. Haigh, “Molecular discrimination of garfish *Hyporhamphus* (Beloniformes) larvae in southern Australian waters,” *Marine Biotechnology*, vol. 3, no. 6, pp. 509–514, 2001.
- [9] J. Zhang, L. Huang, and H. Huo, “Larval identification of *Lutjanus* Bloch in Nansha coral reefs by AFLP molecular method,” *Journal of Experimental Marine Biology and Ecology*, vol. 298, no. 1, pp. 3–20, 2004.
- [10] G. Comi, L. Iacumin, K. Rantsiou, C. Cantoni, and L. Coccolin, “Molecular methods for the differentiation of species used in



- production of cod-fish can detect commercial frauds,” *Food Control*, vol. 16, no. 1, pp. 37–42, 2005.
- [11] P. D. N. Hebert, A. Cywinska, S. L. Ball, and J. R. DeWaard, “Biological identifications through DNA barcodes,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 270, no. 1512, pp. 313–321, 2003.
- [12] P. D. N. Hebert, E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs, “Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fuligator*,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 41, pp. 14812–14817, 2004.
- [13] M. Hajibabaei, D. H. Janzen, J. M. Burns, W. Hallwachs, and P. D. N. Hebert, “DNA barcodes distinguish species of tropical Lepidoptera,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 4, pp. 968–971, 2006.
- [14] D. M. Lambert, A. Baker, L. Huynen, O. Haddrath, P. D. N. Hebert, and C. D. Millar, “Is a large-scale DNA-based inventory of ancient life possible?” *Journal of Heredity*, vol. 96, no. 3, pp. 279–284, 2005.
- [15] M. H. Greenstone, D. L. Rowley, U. Heimbach, J. G. Lundgren, R. S. Pfannenstiel, and S. A. Rehner, “Barcoding generalist predators by polymerase chain reaction: carabids and spiders,” *Molecular Ecology*, vol. 14, no. 10, pp. 3247–3266, 2005.
- [16] R. D. Ward, T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert, “DNA barcoding Australia’s fish species,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1462, pp. 1847–1857, 2005.
- [17] G. G. Pegg, B. Sinclair, L. Briskey, and W. J. Aspden, “MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia,” *Scientia Marina*, vol. 70, no. 2, pp. 7–12, 2006.
- [18] N. Hubert, R. Hanner, E. Holm et al., “Identifying Canadian freshwater fishes through DNA barcodes,” *PLoS ONE*, vol. 3, no. 6, article e2490, 2008.
- [19] M. Vences, M. Thomas, R. M. Bonett, and D. R. Vieites, “Deciphering amphibian diversity through DNA barcoding: chances and challenges,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1462, pp. 1859–1868, 2005.
- [20] M. Vences, M. Thomas, A. van der Meijden, Y. Chiari, and D. R. Vieites, “Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians,” *Frontiers in Zoology*, vol. 2, article 5, 2005.
- [21] M. Aliabadian, M. Kaboli, V. Nijman, and M. Vences, “Molecular identification of birds: performance of distance-based DNA barcoding in three genes to delimit parapatric species,” *PLoS ONE*, vol. 4, no. 1, article e4119, 2009.
- [22] V. Nijman and M. Aliabadian, “Performance of distance-based DNA barcoding in the molecular identification of Primates,” *Comptes Rendus—Biologies*, vol. 333, no. 1, pp. 11–16, 2010.
- [23] S. J. Scheffer, M. L. Lewis, and R. C. Joshi, “DNA barcoding applied to invasive leafminers (Diptera: Agromyzidae) in the Philippines,” *Annals of the Entomological Society of America*, vol. 99, no. 2, pp. 204–210, 2006.
- [24] K. W. Will and D. Rubinoff, “Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification,” *Cladistics*, vol. 20, no. 1, pp. 47–55, 2004.
- [25] R. DeSalle, M. G. Egan, and M. Siddall, “The unholy trinity: taxonomy, species delimitation and DNA barcoding,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1462, pp. 1905–1916, 2005.
- [26] P. H. Barber, S. R. Palumbi, M. V. Erdmann, and M. K. Moosa, “Biogeography: a marine Wallace’s line?” *Nature*, vol. 406, no. 6797, pp. 692–693, 2000.
- [27] N. V. Ivanova, J. R. DeWaard, and P. D. N. Hebert, “An inexpensive, automation-friendly protocol for recovering high-quality DNA,” *Molecular Ecology Notes*, vol. 6, no. 4, pp. 998–1002, 2006.
- [28] N. V. Ivanova, T. S. Zemlak, R. H. Hanner, and P. D. N. Hebert, “Universal primer cocktails for fish DNA barcoding,” *Molecular Ecology Notes*, vol. 7, no. 4, pp. 544–548, 2007.
- [29] M. Kimura, “A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences,” *Journal of Molecular Evolution*, vol. 16, no. 2, pp. 111–120, 1980.
- [30] S. Kumar, K. Tamura, and M. Nei, “MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment,” *Briefings in Bioinformatics*, vol. 5, no. 2, pp. 150–163, 2004.
- [31] J. W. Tuckey, *Exploratory Data Analysis*, Addison–Wesley, Boston, Mass, USA, 1977.
- [32] J. J. Wiens, *Phylogenetic Analysis of Morphological Data*, Smithsonian Institution Press, Washington, DC, USA, 2000.
- [33] J. Köhler, K. Munn, A. Rüegg, A. Skusa, and B. Smith, “Quality control for terms and definitions in ontologies and taxonomies,” *BMC Bioinformatics*, vol. 7, article 212, 2006.
- [34] M. Pfenninger, M. Cordellier, and B. Streit, “Comparing the efficacy of morphologic and DNA-based taxonomy in the freshwater gastropod genus *Radix* (Basommatophora, Pulmonata),” *BMC Evolutionary Biology*, vol. 6, article 100, 2006.
- [35] D. J. Funk and K. E. Omland, “Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA,” *Annual Review of Ecology, Evolution, and Systematics*, vol. 34, pp. 397–423, 2003.
- [36] J. Arnold, “Cytonuclear disequilibria in hybrid zones,” *Annual Review of Ecology and Systematics*, vol. 24, pp. 521–554, 1993.
- [37] T. G. Barraclough and S. Nee, “Phylogenetics and speciation,” *Trends in Ecology and Evolution*, vol. 16, no. 7, pp. 391–399, 2001.
- [38] P. Alexandrino, R. Faria, D. Linhares et al., “Interspecific differentiation and intraspecific substructure in two closely related clupeids with extensive hybridization, *Alosa alosa* and *Alosa fallax*,” *Journal of Fish Biology*, vol. 69, pp. 242–259, 2006.
- [39] J. S. Nelson, R. J. Hoddell, L. M. Chou, W. K. Chan, and V. P. E. Phang, “Phylogeographic structure of false clownfish, *Amphiprion ocellaris*, explained by sea level changes on the Sunda shelf,” *Marine Biology*, vol. 137, no. 4, pp. 727–736, 2000.
- [40] S. F. Chenoweth and J. M. Hughes, “Genetic population structure of the catadromous perciform: *Macquaria novemaculeata* (Percichthyidae),” *Journal of Fish Biology*, vol. 50, no. 4, pp. 721–733, 1997.
- [41] C. L. Dudgeon, N. Gust, and D. Blair, “No apparent genetic basis to demographic differences in scarid fishes across continental shelf of the great barrier reef,” *Marine Biology*, vol. 137, no. 5–6, pp. 1059–1066, 2000.
- [42] G. Bernardi, S. J. Holbrook, and R. J. Schmitt, “Gene flow at three spatial scales in a coral reef fish, the three-spot dascyllus, *Dascyllus trimaculatus*,” *Marine Biology*, vol. 138, no. 3, pp. 457–465, 2001.
- [43] S. R. Palumbi, “Genetic divergence, reproductive isolation, and marine speciation,” *Annual Review of Ecology and Systematics*, vol. 25, pp. 547–572, 1994.

- [44] M. E. Hellberg, R. S. Burton, J. E. Neigel, and S. R. Palumbi, "Genetic assessment of connectivity among marine populations," *Bulletin of Marine Science*, vol. 70, no. 1, pp. 273–290, 2002.
- [45] T. Lefébure, C. J. Douady, M. Gouy, and J. Gibert, "Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation," *Molecular Phylogenetics and Evolution*, vol. 40, no. 2, pp. 435–447, 2006.
- [46] M. C. Ebach and C. Holdrege, "DNA barcoding is no substitute for taxonomy," *Nature*, vol. 434, no. 7034, p. 697, 2005.
- [47] C. C. Nice and A. M. Shapiro, "Patterns of morphological, biochemical, and molecular evolution in the *Oeneis chryxus* complex (Lepidoptera: Satyridae): a test of historical biogeographical hypotheses," *Molecular Phylogenetics and Evolution*, vol. 20, no. 1, pp. 111–123, 2001.
- [48] E. Zuckerkandl and L. Pauling, "Evolutionary divergence and convergence in proteins," in *Evolving Genes and Proteins*, V. Bryson and H. J. Vogel, Eds., pp. 97–166, Academic Press, New York, NY, USA, 1965.