

Molecular Approach to the Identification of Fish in the South China Sea

Junbin Zhang^{1,2*}, Robert Hanner²

1 Department of Aquaculture & Life Science, Shanghai Ocean University, Shanghai, China, **2** Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada

Abstract

Background: DNA barcoding is one means of establishing a rapid, accurate, and cost-effective system for the identification of species. It involves the use of short, standard gene targets to create sequence profiles of known species against sequences of unknowns that can be matched and subsequently identified. The Fish Barcode of Life (FISH-BOL) campaign has the primary goal of gathering DNA barcode records for all the world's fish species. As a contribution to FISH-BOL, we examined the degree to which DNA barcoding can discriminate marine fishes from the South China Sea.

Methodology/Principal Findings: DNA barcodes of cytochrome oxidase subunit I (COI) were characterized using 1336 specimens that belong to 242 species fishes from the South China Sea. All specimen provenance data (including digital specimen images and geospatial coordinates of collection localities) and collateral sequence information were assembled using Barcode of Life Data System (BOLD; www.barcodinglife.org). Small intraspecific and large interspecific differences create distinct genetic boundaries among most species. In addition, the efficiency of two mitochondrial genes, 16S rRNA (16S) and cytochrome *b* (*cytb*), and one nuclear ribosomal gene, 18S rRNA (18S), was also evaluated for a few select groups of species.

Conclusions/Significance: The present study provides evidence for the effectiveness of DNA barcoding as a tool for monitoring marine biodiversity. Open access data of fishes from the South China Sea can benefit relative applications in ecology and taxonomy.

Citation: Zhang J, Hanner R (2012) Molecular Approach to the Identification of Fish in the South China Sea. PLoS ONE 7(2): e30621. doi:10.1371/journal.pone.0030621

Editor: Indra Neil Sarkar, University of Vermont, United States of America

Received: February 26, 2011; **Accepted:** December 22, 2011; **Published:** February 17, 2012

Copyright: © 2012 Zhang, Hanner. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported through funding to the Canadian Barcode of Life Network from Genome Canada (through the Ontario Genomics Institute) and by Chinese National Funding awards 40776089 and U0633007, which supported specimen collection. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jb-zhang@shou.edu.cn

Introduction

Fishes show an astonishing diversity of shapes, sizes, and colors. The delimitation and recognition of fish species is not only of interest for taxonomy and systematists, but it is also a requirement in studies of natural history and ecology, fishery management, tracking the dispersal patterns of eggs and larvae, estimations of recruitment and spawn areas, and authentication of food products [1–2]. Fish identification is traditionally based on morphological features. However, due to high diversity and morphological plasticity, in many cases, fish and their diverse developmental stages are difficult to identify by using morphological characteristics alone [2]. DNA-based identification techniques have been developed and proven to be analytically powerful [3–5]. As a standardized and universal method, DNA barcoding identification systems have been widely advocated to identify species and uncover biological diversity in these years [6–7]. For many animal taxa, sequence divergences within the 5' region of the mitochondrial cytochrome oxidase subunit I (COI) gene are generally much greater between species than within species. This in turn suggests that the approach is extensively applicable among phylogenetically distant animal groups [8–14]. Many studies have shown that

intraspecific variation of COI barcodes is generally pretty small and clearly discriminable from interspecific variation [15–24].

The South China Sea lies within the Indo-West Pacific marine biogeographic province, which has long been recognized as the global center of marine tropical biodiversity [25]. In addition to temperate species, there are many coral fish living in the South China Sea. The most striking feature of these marine fish is their diversity, both in terms of number of species and in the range of morphologies [26]. In the present study, more than 1,300 specimens from the South China Sea were sequenced for COI barcodes. DNA barcode data were then integrated with the relevant taxonomical and ecological information in two projects, Fishes from the South China Sea (FSCS) and Coral Fishes from the South China Sea (CFCS), in the Barcode of Life Data System (BOLD).

Recently, some other mitochondrial genes or nuclear ribosomal DNA fragments, have been proposed as alternatives for species identification [5,27]. Most studies focus on narrow-range taxa, but only a few have systematically compared the utility of different molecular markers in species identification. Herein, we also used samples from a few select groups of species to test three other different molecular markers—mitochondrial cytochrome *b* (*cytb*),

16S rRNA gene (16S), and nuclear ribosomal 18S rRNA gene (18S)—with respect to their ability to identify fish species.

Materials and Methods

Ethics

Ethical approval was not required for this study because no endangered fish were involved. However, specimen collection and maintenance were performed in strict accordance with the recommendations of Animal Care Quality Assurance in China.

Specimen collection and DNA extraction

Fish samples were collected from more than 40 locations in the South China Sea (Fig. 1, Table S1). Voucher specimens were deposited in the Marine Biodiversity Collection of South China Sea, South China Sea Institute of Oceanography, Chinese Academy of Sciences. All specimens were preserved in 70% ethanol. Tissue samples were dissected from the body muscle, and genomic DNA was extracted according to the standard Barcode of Life protocol [28].

PCR and DNA sequencing

Fragments of the 5' region of mitochondrial COI gene were amplified using C_FishF1t1/C_FishR1t1 primer cocktails [29]. The primer combination C_FishF1t1 contained two primers (FishF2_t1/VF2_t1), and C_FishR1t1 also contained two primers (FishR2_t1/FR1d_t1). These primers are described in Table S2. PCR reactions were carried out in 96-well plates using Mastercycler® Eppendorf gradient thermal cyclers (Brinkmann Instruments, Inc.). The reaction mixture of 825 µl water, 125 µl 10× buffer, 62.5 µl MgCl₂ (25 mM), 6.25 µl dNTP (10 mM), 6.25 µl of each primer (0.01 mM), and 6.25 µl Taq DNA polymerase (5 U/µl) was prepared for each plate. Each well contained 10.5 µl of mixture and 2 µl genomic DNA. Thermocycling comprised an initial step of 2 min at 95°C and 35 cycles of 30 s at 94°C, 40 s 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). Each chosen PCR product was sequenced bi-directionally with the primers M13F (5'-TGTAACGACGGCCAGT-3') and M13R (5'-CAGGAAA-CAGCTATGAC-3') using the BigDye® Terminator v.3.1 Cycle Sequencing Kit (PE Biosystems, Inc.). Thermocycling conditions were as follows: An initial step of 2 min at 96°C and 30 cycles of 30 s at 96°C, 15 s at 55°C, and 4 min at 60°C. Final PCR products were directly sequenced using an ABI 3730 capillary sequencer according to manufacturer's instructions.

For specimens that failed to yield amplification products 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) (Table S2). The PCR program consisted of an initial denaturation at 94°C for 1 min, five cycles of 94°C for 30 s, annealing at 50°C for 40 s, and extension at 72°C for 1 min, followed by 30 cycles of 94°C for 30 s, 54°C for 40 s, and 72°C for 1 min, with a final extension at 72°C for 10 min. All other procedures were performed as given above.

Specimen data such as images, collection information, museum accession numbers, and sequence trace files were assembled in BOLD in accordance with the BARCODE data standard as specified by the Consortium for the Barcode of Life in collaboration with the International Nucleotide Sequence Database Collaboration (INSDC) [30,31]. Sequences were submitted to GenBank using the NCBI Barcode Submission Tool, where they were subsequently annotated with the reserved keyword BARCODE.

In addition to the COI barcode region, two DNA fragments, one of mitochondrial 16S and one of *cytb*, and one DNA fragment of nuclear ribosomal 18S were screened as potential species markers in 282 specimens from 52 species. Primers utilized in this study are listed in Table S2. Each PCR reaction mixture

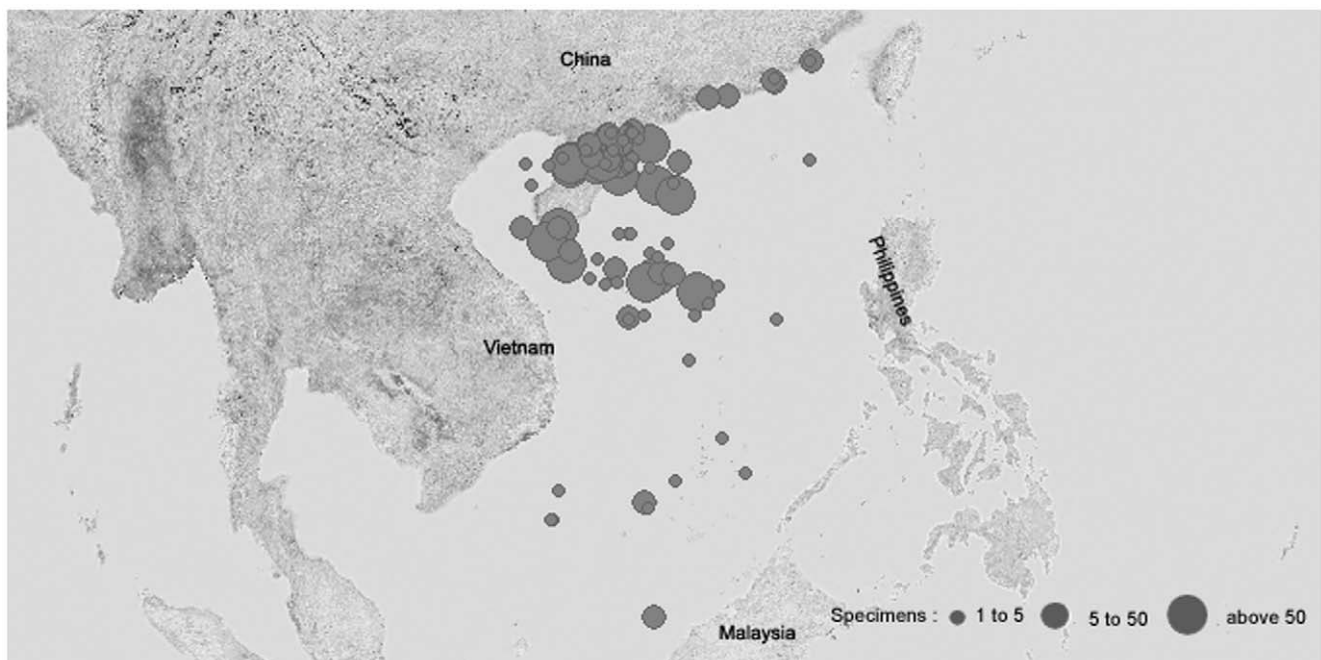


Figure 1. Map showing the sampling localities for fish from the South China Sea.
doi:10.1371/journal.pone.0030621.g001

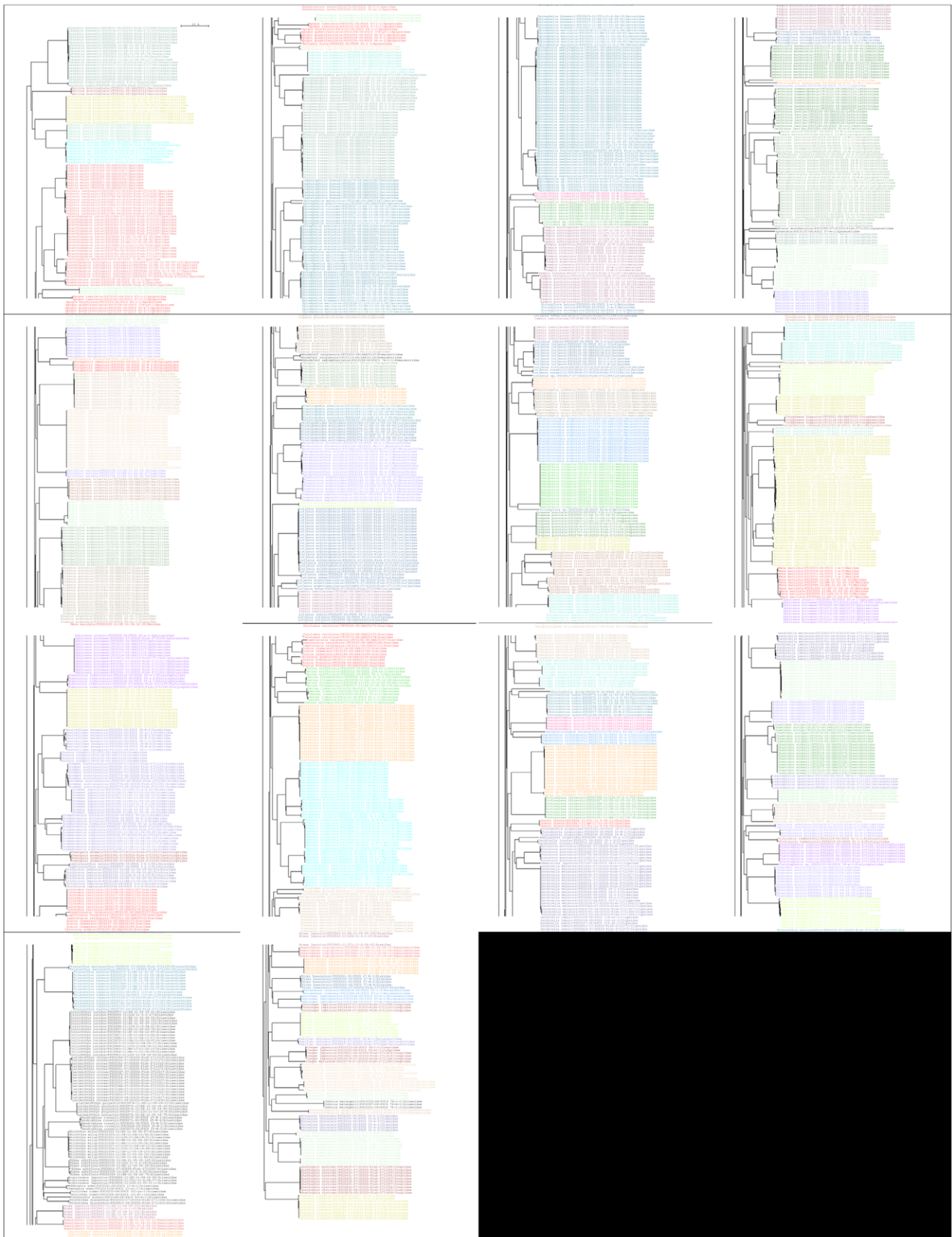


Figure 2. Neighbor-joining (NJ) tree based on COI barcodes from fishes of the South China Sea. Scale: 5% K2P distance. The specimen ID is annotated in each sequence.
doi:10.1371/journal.pone.0030621.g002

Table 1. Summary of K2P genetic divergences at different taxonomic levels.

Comparisons within	Taxa	Number of comparisons	Mean	Minimum	Maximum	SE*
species	199	5814	0.18	0	2.51	0.01
genus	141	5958	13.55	0	25.35	0.07
family	80	11126	19.65	6.99	35.71	0.051
order	22	375239	24.05	13.96	39.58	0.01
class	2	429246	24.91	14.74	40.58	0.01

*standard error.

Data are from 1336 sequences from 242 species and 159 genera.

doi:10.1371/journal.pone.0030621.t001

contained of 16.7 μ l water, 2.5 μ l 10 \times buffer, 2.0 μ l MgCl₂ (25 mM), 1 μ l dNTPs (10 mM), 0.5 μ l each primer (0.01 mM), 0.2 μ l Taq DNA polymerase (5 U/ μ l) and 1.0 μ l template DNA. PCR amplifications were performed with the following conditions: 35 cycles of denaturation at 94°C for 45 s, annealing at 52–62°C (depending on the primer combination) for 50 s, and extension at 72°C for 1 min, with an initial denaturation at 94°C for 2 min and final extension at 72°C for 5 min. Amplified products were visualized in 1% agarose gel, and purified products were directly sequenced on an Applied Biosystems 3730 sequencer using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems, Inc.). Sequencing primers were the same as those listed above for PCR. All sequencing reactions were performed according to the manufacturer's instructions.

Data analyses

DNA sequences were aligned with SeqScape v.2.1.1 software (Applied Biosystems, Inc.). Mitochondrial COI and *cytb* sequences were translated into amino acids in order to exclude sequencing errors and to avoid the inclusion of pseudogene sequences in the datasets. Sequence divergences were calculated using the Kimura 2 parameter (K2P) distance [32]. This system usually makes a suitable metric model when genetic distances are low [33]. An unrooted NJ tree based on K2P distances was created using MEGA software (version 3.1) [34].

The following categories of K2P distances were calculated: intraspecific distances, interspecific values within the same genus, and interspecific values between different genera within the same family. These values were plotted using the boxplot representation of R. Boxplots in SPSS 11.5 software (SPSS Inc., Chicago, IL, U.S.) [35]. Separate boxplots were constructed only for families containing specimens from 2 or more genera in order to compare among taxonomic categories. 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) are described in the boxplots.

Results

COI DNA barcoding

A total of 1336 bidirectional COI sequences belonging to 242 species were obtained (GenBank accession numbers, taxonomic data and museum numbers listed in Table S1). 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.

The mean intraspecific K2P (Kimura two-parameter) distance was 0.18%. The distance increased sharply to 13.55% among

individuals of different congeneric species. Apart from *Pampus*, all other COI sequences formed species clusters [Fig. 2]. Barcode divergences of 1% were used as filters to perform comparisons between units that were identified morphologically; the criterion was met in all cases except *Upeneus sulphureus*, *Siganus guttatus*, *Alepes djedaba*, *Acentrogobius caninus*, *Hyporhamphus limbatus*, *Gymnothorax reevesii*, *Kumococcus rodericensis*, *Mene maculata*, *Terapon jarbua*, *Zebrias quagga*, *Pennahia anea*, and *Mugil cephalus*. For these the barcode divergences reached maximum value of 2.51%, and 98.43% (5723 out of 5814) of pairwise genetic distances within species were below 1%. Overall, the average of interspecific distances among congeneric species was over 70-fold higher than that of intraspecific distances. For higher taxonomic ranks (family, order, and class), mean pairwise genetic distances increased gradually, reaching 19.65%, 24.05%, and 24.91%, respectively (Table 1). Interspecific genetic distances below 5% were found only among pairwise comparisons within genera and not at high taxonomic levels such as family or order. The steep increase in genetic variation at the generic level and the smoothness of the rise at high taxonomic levels was observed. This indicates profound differences at species boundaries under the frame of COI divergence (Table 1 and Fig. 3). The distribution of the nearest-neighbor distance (NND), namely the minimum of genetic variation between a

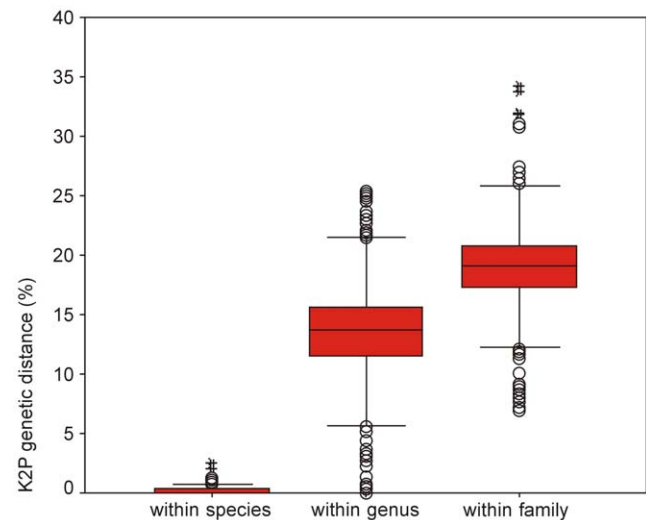


Figure 3. The distribution of K2P distances at COI sequences within species, genus and family respectively. IQR: interval into which the central 50% of the data fall. Black bar in the box indicates the median. Circles indicate mild outliers and asterisks indicate extreme outliers. Extreme outliers are discussed in the text.

doi:10.1371/journal.pone.0030621.g003

species and its closest relative, revealed that only 3.31% of NNDs (8 cases) were lower than 1% (Fig. 4). Fish speciation has many causes, and the rate of mitochondrial COI differentiation during evolution is not equal for all fishes [27]. The distribution of interspecific K2P genetic distances of COI gene within genera at the family level was obviously different (Table 2). Wide fluctuations were observed in values of the interspecific divergence within genera. In the genus *Gerres*, the interspecific distance reached 25.35%, but in the genera *Scomber*, *Thamnaconus*, *Pterois*, *Cololabis*, *Etmopterus*, *Pampus*, and *Plectropomus*, most genetic variations within the genus were below 5%.

Genetic analyses of other markers

A high level of sequence variations for *cytb* makes it difficult to design universal primers for these fish. Thirteen primers were designed for *cytb* (Table S2), but fewer than half of the samples were amplified successfully. For the 282 selected specimens from 52 species, a data set of 281 mitochondrial 16S (521–561 bp; accession numbers JN211430–JN211710), 124 *cytb* (832 bp; accession numbers JN211987–JN212110), and 276 nuclear ribosomal 18S (449–459 bp; accession numbers JN211711–JN211986) sequences were ultimately obtained. Many insertions

and deletions were found in 16S and 18S. While sequence errors could be detected for *cytb* by translating into amino acids, the non-coding regions of 16S and 18S could not. The average intraspecific variation was 0.78 for *cytb* and 0.27 for 16S. Intraspecific K2P distances of 18S were low (the average was only 0.16), and 18S sequences were conserved across a broad range of taxa (Table 3 and Fig. 5). In some congeneric species, no genetic variations were observed. These included *Epinephelus coioides* and *Epinephelus maculatus* (Fig. 6). Due to its high sequence conservation, distance-based inference may not be appropriate for 18S analysis as an approach to species assignment. The character-based method advocated by Sarkar et al. may be a suitable alternative [36]. In COI analysis based on the criterion of genetic distance, deep intraspecific divergences were observed in *Mene maculata* and *Terapon jarbua*, but unique type was characterized for each species based on the sequence analysis of 18S (Fig. 6). Exploring several gene regions for species markers and choosing a gene region and an appropriate measure for species identification can balance the potential for two types of errors: (1) mistreating individual variation for species level variation by using a relatively variable gene region; or (2) failing to identify true species differences, by using a conserved gene region to recover sufficient variation [37].

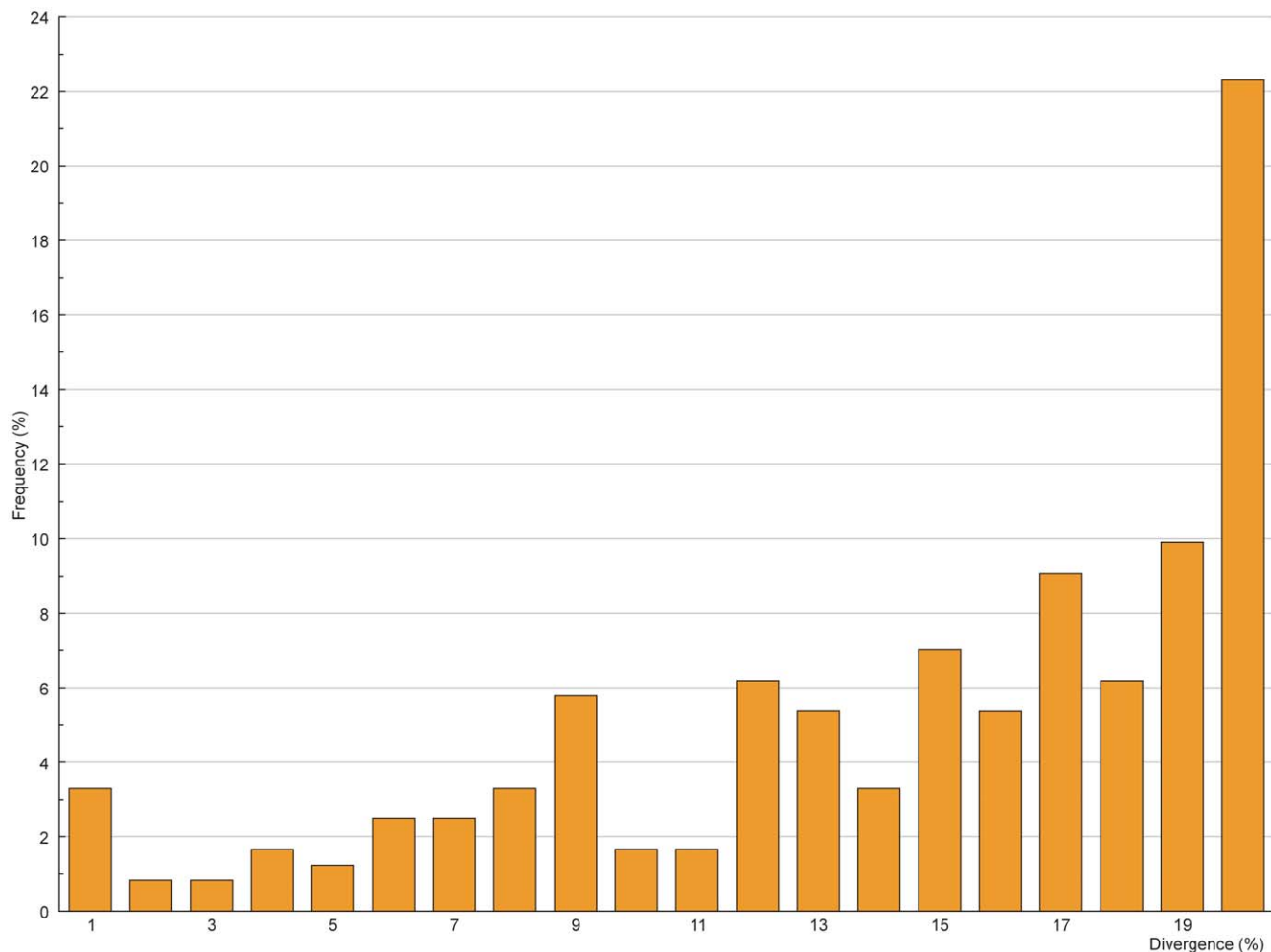


Figure 4. Distribution of the genetic distances to the nearest-neighbor. The analysis is based on all the comparisons of COI barcodes from this study.

doi:10.1371/journal.pone.0030621.g004

Table 2. Distribution of interspecific K2P genetic distances of COI gene within genus at the family level.

Order	Family	<5%	5–10%	>10%
Anguilliformes	Congridae			5
Anguilliformes	Muraenidae	3		24
Aulopiformes	Synodontidae			6
Beloniformes	Belonidae			11
	Hemiramphidae			36
	Scomberesocidae	12		
Clupeiformes	Chirocentridae			8
	Clupeidae			156
	Engraulidae			41
Mugiliformes	Mugilidae			19
Perciformes	Apogonidae			14
	Carangidae			86
	Centropomidae		32	
	Chaetodontidae	52	37	
	Gerreidae			59
	Haemulidae			30
	Labridae			7
	Leiognathidae		15	
	Lethrinidae			56
	Lutjanidae		15	560
	Mullidae			220
	Pomacentridae		2	
	Priacanthidae			28
	Sciaenidae			80
	Scombridae	72		48
	Serranidae		237	3217
	Siganidae			68
Sillaginidae			32	
Sparidae			111	
Sphyraenidae			13	
Stromateidae	24	19	236	
Terapontidae			8	
Pleuronectiformes	Cynoglossidae		9	81
Rajiformes	Dasyatidae			38
Scorpaeniformes	Platycephalidae	20		34
Squaliformes	Etmopteridae	16		
Tetraodontiformes	Monacanthidae	18		
	Total	217	462	5279

Values of zero have been left blank.
doi:10.1371/journal.pone.0030621.t002

Discussion

The ideal DNA barcoding should be robust, with conserved priming sites and reliable DNA amplifications and sequencing, and the DNA fragment sequenced should be nearly identical among individuals of the same species, but differentiative between species [38]. Therefore, we hope that DNA sequences exhibit high levels of conservation within the species and modest levels of genetic variability between different species. [39]. If the gene evolves too quickly, genetic variation would tend to be saturated

Table 3. Summary of K2P genetic p-distances (%) within different taxonomic levels.

Comparisons within	16S		Cyt <i>b</i>		18S	
	Mean	SE	Mean	SE	Mean	SE
species	0.27	0.05	0.78	0.11	0.16	0.02
genus	2.38	0.36	10.52	0.32	0.54	0.29
family	4.39	0.42	15.24	0.61	4.21	0.08
order	11.82	0.85	-	-	6.72	0.17

Values are calculated from DNA partial sequences of mitochondrial 16S rRNA (16S; n=281), cytochrome *b* (cyt *b*; n=124), and nuclear ribosomal 18S rRNA (18S; n=276).

doi:10.1371/journal.pone.0030621.t003

at lower taxonomic groups. However, if it evolves too slowly, some closely allied species may not be differentiated. In other words, the high level of sequence conservation across a wide range of taxa can underestimate species diversity [40]. In this study, interspecific variations within the genera and families were close for 16S and 18S (Table 3 and Fig. 5). The presence of insertions and deletions in 16S and 18S can lead to errors in sequence alignment [41]. Compared to protein-coded COI and *cytb*, the design of *cytb* primers is surprisingly difficult given that COI is usually more conserved than *cytb*. Based on the comprehensive analyses given above, the results show that the COI barcode region is a more suitable species marker across wide-range taxa.

Other DNA markers can provide assistance to species identification in cases where COI is lack of high resolving power. While DNA barcoding provides taxonomic identification for a given specimen, accuracy depends on whether there is an exact or nearly match to that species in the database. It is desirable that COI sequences representing each taxon in the reference database can cover the major part of the existing diversity, otherwise in the interrogation of BOLD, identification difficulties would arise when the unknown specimens come from a currently under-described

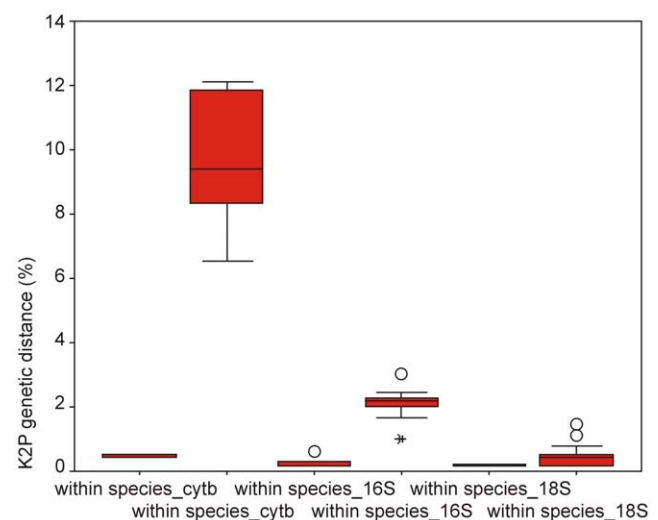


Figure 5. The distribution of intra- and inter-specific genetic divergences for *cytb*, 16S, and 18S. K2P genetic distances within species and genus for partial sequences from mitochondrial *cyt b*, 16S, and nuclear ribosomal 18S genes of fish from the South China Sea.
doi:10.1371/journal.pone.0030621.g005

	1	1	1	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3
	6	1	2	2	6	6	7	9	9	9	2	2	3	5	0	1	1	1	1	3	3	3	3	3	
	4	2	1	8	0	8	3	3	4	9	0	6	1	3	3	1	5	8	9	4	0	8	9	9	
<i>E. coioides</i> FSCS924-11	A	A	C	T	C	A	A	C	T	T	C	C	C	C	C	T	A	C	G	T	G	-	T		
<i>E. coioides</i> FSCS925-11	
<i>E. coioides</i> FSCS928-11	
<i>E. maculatus</i> FSCS933-11	
<i>E. maculatus</i> FSCS935-11	
<i>E. maculatus</i> FSCS936-11	
<i>E. amblycephalus</i> FSCS910-11	T	
<i>E. amblycephalus</i> FSCS913-11	T	
<i>E. amblycephalus</i> FSCS916-11	T	
<i>M. maculata</i> FSCS994-11	C	C	T	A	.	.	
<i>M. maculata</i> FSCS995-11	C	C	T	A	.	.	
<i>M. maculata</i> FSCS996-11	C	C	T	A	.	.	
<i>M. maculata</i> FSCS998-11	C	C	T	A	.	.	
<i>M. maculata</i> FSCS999-11	C	C	T	A	.	.	
<i>T. jarbua</i> FSCS120-06	T	G	A	C	T	G	G	T	.	C	.	T	T	T	T	A	.	T	C	C	T	-	G		
<i>T. jarbua</i> FSCS122-06	T	G	A	C	T	G	G	T	.	C	.	T	T	T	T	A	.	T	C	C	T	-	G		
<i>T. jarbua</i> FSCS123-06	T	G	A	C	T	G	G	T	.	C	.	T	T	T	T	A	.	T	C	C	T	-	G		
<i>T. jarbua</i> FSCS293-06	T	G	A	C	T	G	G	T	.	C	.	T	T	T	T	A	.	T	C	C	T	-	G		

Figure 6. Sequence analyses of 18S for fish identification. Diagnostic sites in 18S for *Epinephelus coioides*, *Epinephelus maculatus*, *Epinephelus amblycephalus*, *Mene maculata*, *Terapon jarbua*, and *Zebrias quagga* as examples of the character-based method.
doi:10.1371/journal.pone.0030621.g006

part of biodiversity [42]. In case of low resolution from the COI gene alone, the combination of other molecular markers such as *cytb*, 16S, and 18S can help solve this problem. For example, intraspecific variations of the COI gene in *Mene maculata* and *Terapon jarbua* were greater than the average of most intraspecific values, which imply possible overlaps with close related species if the sampling size is augmented continuously. In such cases, the sequence analysis of 18S sequences or other markers could help resolve this overlap should it occur.

Geographical structure, if ignored, can blur and distort species delineation [27]. Biological mechanisms, water dynamics, and even historical events may affect the deep genetic structure of marine populations [43]. Many explanations of genetic population structure on local and regional scales involve behaviors such as the adoption of pelagic early life stages and movement over broad geographic ranges. These factors are theoretically associated with gene flow. For marine fish, there is generally a lack of genetic differentiation within species on macrogeographic scales [44–47]. In this study, for many species, intraspecific genetic variations were near or equal to zero. However, some pairwise K2P distances of more than 1% were observed. Deep intraspecific genetic divergences were observed in species displaying restricted migratory behaviors or other biological mechanisms that would limit gene flow among individuals [48,49]. *Siganus guttatus*, *Alepes djedaba*, *Scomber japonicus*, *Hyporhamphus limbatus*, *Terapon jarbua*, and *Pennahia anea* are coastal marine fish that reproduce in estuaries and bays and do not undertake large-scale migratory movements. The relevance of the reference DNA barcode database depends on the exhaustiveness of intra-taxon sampling, so the global participation and cooperation is indispensable for DNA barcoding projects.

The combination of morphological and molecular characteristics can bridge the gap between morphological taxonomy and the DNA barcoding approach [37]. This idea has been embodied in the establishment of BOLD. DNA sequences in BOLD are derived from voucher specimens preserved in museums all around the world. Specimen data such as photo images and collection information are linked with each sequence. One can solve any problems concerning morphological identification by searching the relevant database or sending inquiries to confirm voucher specimens. The taxonomy of Leiognathidae species has changed drastically as a result of revisions carried out in recent years [50]. Several taxonomic designations of species used in the literature have been recognized as dubious identifications [51]. For example, *Nuchequula nuchalis* is misidentified as *Leiognathus nuchalis* [52] or *Leiognathus blochii* [53], and *Equulites leuciscus* is misidentified as *Leiognathus leuciscus* [53]. In this study, all genetic distances between *Nuchequula nuchalis* and *Equulites leuciscus* are over 15.80% [Fig. 7], and the value is greater than the average (13.55%) within genus. The big divergence among individuals of the two species supports the current taxonomy about Leiognathidae in which they should be classified into different genera [51]. In the genus *Pampus*, there are overlaps between intraspecific and interspecific genetic variations [Fig. 2]. Due to morphological similarities in *Pampus*, there is great confusion regarding the relative nomenclature [54–57]. *P. cinereus* is regarded by Parin and Piotrovsky as a synonym of *P. argenteus* based on morphological characteristics [57]. In the present study, *P. cinereus* and *P. argenteus* show small genetic variations and overlap in the NJ tree [Fig. 2], and our results support the idea that the nomenclature of *Pampus cinereus* may be removed as the synonym of *Pampus argenteus* in the FISHBASE. The results of DNA barcoding can also provide clues to the

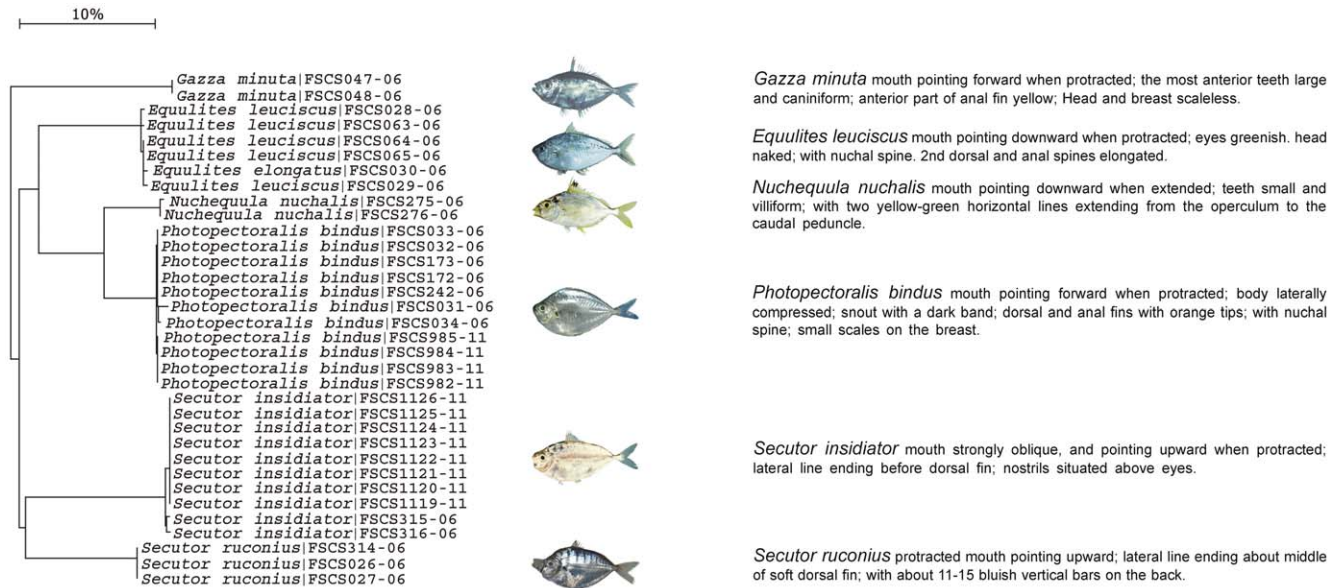


Figure 7. Neighbour joining tree (K2P) for Leiognathidae samples. Branches with specimen ID-number from BOLD and species name. doi:10.1371/journal.pone.0030621.g007

discovery of sample misidentification. One specimen of *Thrissa kammalensis*, which was collected off the west coast of the South China Sea, showed an average genetic divergence of 9.25% from other individuals of *Thrissa kammalensis*. However, its sequence was identical to those of *Thrissa setirostris*. The identification of this specimen merits suspicion because the value 9.25% greatly exceeds the average intraspecific genetic range. We checked the voucher specimen and found that this particular case had been misclassified. Species identification generally requires the collection of a large number of individuals, and occasional instances of misclassification are perhaps inevitable. Voucher specimens must be preserved in good condition for later collaborations and deposited for posterity in longstanding, legitimate collections dedicated to the storage of such materials [58]. Moreover, this example suggests that DNA barcoding can detect cases of morphological misclassification. The Fish Barcode of Life (FISH-BOL) campaign has the primary goal of gathering DNA barcode records for all of the world's fish. Standard reference DNA sequences amplified from expertly identified morphological voucher specimens can be used to better characterize and broadly identify species [10,59].

One of the key concerns raised against DNA barcoding is that nuclear mitochondrial pseudogenes (numts) may misestimate the number of unique species [60,61]. Actually, such problems were taken into account at the beginning of the DNA barcoding project [62]. Generally, submitted sequences are evaluated for suspicious numts in Barcode of Life Data Systems (BOLD, www.barcodinglife.org) if indels or stop codons are found. It seems possible that some numts may be of the expected length without any in-frame stop codons and therefore may not be readily distinguishable from the orthologous mtDNA [60]. Definite diagnosis is confirmed only by large numbers of sequence comparisons within and between species. We can set up a sub-database for numts in BOLD. After the abundant influx of the relevant data, the misidentification rate will dramatically decrease. In this study, over 1,000 specimens were amplified using universal primers, and only 4 numts were obtained, all of them in *Satyriothys amiscus*. Orthologous mtDNAs were success-

fully amplified only by increasing the annealing temperature by 2°C. The number of mitochondrial genomes is greater than that of nuclear genomes, so conserved primers should preferentially amplify mtDNAs over numts. In special cases, several methods have been suggested as means of avoiding numt co-amplification. These include RT-PCR, long PCR, and mtDNA enrichment [63].

Supporting Information

Table S1 Specimen data and GenBank accession numbers of these 1336 sequences (the specimen ID is the number which is given for each specimen in Barcode of Life Database (www.barcodinglife.org) containing all information from two projects, "Fishes from the South China Sea" (FSCS) and "Coral Fishes from the South China Sea" (FSCS). (DOCX)

Table S2 Primers (5'-3') utilized for PCR amplifications and sequencing in this study. (DOCX)

Acknowledgments

This data release paper represents a joint Chinese-Canadian contribution to the global campaign to barcode all fishes, FISH-BOL. We gratefully acknowledge the assistance of Professor Tingbao Yang of Zhongshan University for assistance with morphological identifications, Christa Maitland for assistance in the lab, Eugene Wong for assistance with phylogenetic analyses and Robin Floyd for comments on an earlier version of this manuscript. We thank Paul Hebert and our colleagues at the Canadian Centre for DNA Barcoding and the BOLD analytics staff at the Biodiversity Institute of Ontario for technical support. We also acknowledge Marine Biodiversity Collection of South China Sea, Chinese Academy of Sciences for archiving these voucher specimens.

Author Contributions

Conceived and designed the experiments: RH JBZ. Performed the experiments: JBZ. Analyzed the data: JBZ. Contributed reagents/materials/analysis tools: JBZ RH. Wrote the paper: JBZ RH.

References

- Rasmussen RS, Morrissey MT, Hebert PDN (2009) DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmo*) from North America. *J Agric Food Chem* 57: 8379–8385.
- Victor BC, Hanner R, Shivji M, Hyde J, Caldwell C (2009) Identification of the larval and juvenile stages of the Cubera snapper, *Lutjanus cyanopterus*, using DNA barcoding. *Zootaxa* 2215: 24–36.
- Zhang JB, Huang LM, Huo HQ (2004) Larval identification of *Lutjanus* Bloch in Nansha coral reefs by AFLP molecular method. *J Exp Mar Biol Ecol* 298: 3–20.
- Comi G, Iacumin L, Rantsioua K, Cantoni C, Coccolin L (2005) Molecular methods for the differentiation of species used in production of cod-fish can detect commercial frauds. *Food Control* 16: 37–42.
- Teletchea F (2009) Molecular identification methods of fish species: reassessment and possible applications. *Rev Fish Biol Fish* 19: 265–293.
- Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc Biol Sci B* 270: 313–321.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci USA* 101: 14812–14817.
- Lambert DM, Baker A, Huynen L, Haddrath O, Hebert PDN (2005) Is a large-scale DNA-based inventory of ancient life possible? *J Hered* 96: 279–284.
- Greenstone MH, Rowley DL, Heimbach U, Lundgren JG, Pfannenstiel RS, et al. (2005) Barcoding generalist predators by polymerase chain reaction: carabids and spiders. *Mol Ecol* 14: 3247–3266.
- Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philos Trans Royal Soc B* 360: 1847–1857.
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci USA* 103: 968–971.
- Pegg GG, Sinclair B, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Sci Mar* 10: 7–12.
- Scheffer SJ, Lewis ML, Joshi RC (2006) DNA barcoding Applied to invasive leafminers (Diptera: Agromyzidae) in the Philippines. *Ann Entomol Soc Am* 99: 204–210.
- Neigel J, Domingo A, Stake J (2007) DNA barcoding as a tool for coral reef conservation. *Coral Reefs* 26: 487–499.
- Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, et al. (2008) Identifying Canadian freshwater fishes through DNA barcodes. *PLoS One* 3: e2490.
- Rock J, Costa FO, Walker DI, North AW, Hutchinson WF, et al. (2008) DNA barcodes of fish of the Antarctic Scotia Sea indicate priority groups for taxonomic and systematics focus. *Antarctic Sci* 20: 253–262.
- Swartz ER, Mwalea M, Hanner R (2008) A role for barcoding in the study of African fish diversity and conservation. *S Afr J Sci* 104: 293–298.
- Steinke D, Zemlak TS, Hebert PDN (2009a) Barcoding Nemo: DNA-based identification for the marine ornamental fish trade. *PLoS One* 4: e6300.
- Steinke D, Zemlak TS, Boutilier JA, Hebert PDN (2009b) DNA barcoding of Pacific Canada's fishes. *Mar Biol* 156: 2641–2647.
- Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. *J Fish Biol* 74: 329–356.
- Barbuto M, Galimberti A, Ferri E, Labra M, Malandra R, et al. (2010) DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of “palombo” (*Mustelus* spp.). *Food Res Int* 43: 376–381.
- Kochzius M, Seidel C, Antoniou A, Botla SK, Campo D, et al. (2010) Identifying fishes through DNA barcodes and microarrays. *PLoS One* 5: e12620.
- Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine metazoa. *Ann Rev Mar Sci* 3: 471–508.
- Wong LL, Peatman E, Lu J, Kucuktas H, He S, et al. (2011) DNA barcoding of catfish: species authentication and phylogenetic assessment. *PLoS One* 6: e17812.
- Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2000) A marine Wallace's line? *Nature* 406: 692–693.
- Lucey K (2010) Methods of Speciation in Tropical Reef Fish. *Rollins Undergraduate Res J*. 2 p. Available at: <http://scholarship.rollins.edu/rurj/vol2/iss1/8>. Accessed 2010 Feb 10.
- Frézal L, Leblois R (2008) Four years of DNA barcoding: current advances and prospects. *Infect Genet Eol* 8: 727–736.
- Ivanova NV, deWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high quality DNA. *Mol Ecol Notes* 6: 998–1002.
- Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Mol Ecol Notes* 7: 544–548.
- Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes* 7: 355–364.
- Hanner R (2005) Proposed standards for BARCODE records in INSDC. Database Working Group, Consortium for the Barcode of Life. Available at: http://www.barcoding.si.edu/pdf/dwg_data_standards-Final.pdf.
- Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111–120.
- Nei M, Kumar S (2000) Molecular evolution and phylogenetics. New York: Oxford University Press. pp 333.
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5: 150–163.
- Tuckey JW (1977) Exploratory data analysis. Boston: Addison–Wesley. pp 688.
- Sarkar IN, Joseph WT, Paul JP, David HF, Bernd S, et al. (2002) An automated phylogenetic key for classifying homeoboxes. *Mol Phylogenet Evol* 24: 388–399.
- DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Phil Trans R Soc B* 360: 1905–1916.
- Valentini A, Pompanon F, Taberlet P (2008) DNA barcoding for ecologists. *Trends Ecol Evol* 24: 110–117.
- Ornelas-García CP, Dominguez-Dominguez O, Doadrio I (2008) Evolutionary history of the fish genus *Astyanax* Baird & Girard (1854) (Actinopterygii, Characidae) in Mesoamerica reveals multiple morphological homoplasies. *BMC Evol Biol* 8: e340.
- Piganeau G, Eyre-Walker A, Grimsley N, Moreau H (2011) How and why DNA barcodes underestimate the diversity of microbial eukaryotes. *PLoS One* 6: e16342.
- Doyle JJ, Gaut BS (2000) Evolution of genes and taxa: a primer. *Plant Mol Biol* 42: 1–6.
- Rubinoff D (2006) Utility of mitochondrial DNA barcodes in species conservation. *Cons Biol* 20: 1026–1033.
- Chenoweth SF, Hughes JM (1997) Genetic population structure of the catadromous Perciform: *Macquaria novemaculeata* (Percichthyidae). *J Fish Biol* 50: 721–733.
- Dudgeon CL, Gust N, Blair D (2000) No apparent genetic basis to demographic differences in scarid fishes across continental shelf of the Great Barrier Reef. *Mar Bio* 137: 1059–1066.
- Bernardi G, Holbrook SJ, Schmitt RJ (2001) Gene flow at three spatial scales in a coral reef fish, the three-spot dascyllus, *Dascyllus trimaculatus*. *Mar Bio* 138: 457–465.
- Palumbi SR (1994) Genetic divergence, reproductive isolation and marine speciation. *Ann Rev Ecol Syst* 25: 547–572.
- Hellberg ME, Burton RS, Neigel CN (2002) Genetic assessment of connectivity among marine populations. *Bul Mar Sci* 70: 273–290.
- Planes S, Doherty PJ, Bernardi G (2001) Strong genetic divergence among populations of a marine fish with limited dispersal, *Acanthochromis polyacanthus*, within the Great Barrier Reef and the Coral Sea. *Evolution* 55: 2263–2273.
- Santos S, Hrbek T, Farias IP, Schneider H, Sampaio I (2006) Population genetic structuring of the king weakfish, *Macrondon ancylodon* (Sciaenidae), in Atlantic coastal waters of South America: deep genetic divergence without morphological change. *Mol Ecol* 15: 4361–4373.
- Sparks JS, Dunlap PV, Smith WL (2005) Evolution and diversification of a sexually dimorphic luminescent system in ponyfishes (Teleostei: Leionathidae), including diagnoses of two new genera. *Cladistics* 21: 305–327.
- Chakrabarty P, Sparks JS, Ho HC (2010) Taxonomic review of the ponyfishes (Perciformes: Leionathidae) of Taiwan. *Mar Biodiv* 40: 107–121.
- Chen JTF, Yu MJ (1986) A synopsis of the vertebrates of Taiwan. Taipei: The Commercial Press.
- Shen SC Coastal fishes in Taiwan. Taipei: National Taiwan University Press. pp 180.
- Bernardi G, Holbrook SJ, Schmitt RJ, Crane NL, DeMartini E (2002) Species boundaries, populations and colour morphs in the coral reef three-spot damselfish (*Dascyllus trimaculatus*) species complex. *Proc R Soc Biol Sci B* 269: 599–605.
- Cui ZX, Liu Y, Liu J, Luan WS (2010) Molecular identification of *Pampus* fishes (Perciformes, Stromateidae). *Ichthyol Res* 57: 32–39.
- Liu J, Li CS, Li XS (2002) Studies on Chinese pomfret fishes of the genus *Pampus* (Pisces: Stromateidae). *Stud Mar Sinica* 44: 204–252.
- Parin NV, Piotrovsky AS (2004) Stromateoid fishes (Suborder: Stromateoidei) of the Indian Ocean (species composition, distribution, biology and fisheries). *J Ichthyol* 44: 33–62.
- Ruedas LA, Salazar-Bravo J, Drago JW, Yates TL (2000) The importance of being earnest: what, if anything, constitutes a “specimen examined?” *Mol Phylogenet Evol* 17: 129–132.
- Ekrem T, Willassen E, Stur E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Mol Phylogenet Evol* 43: 530–542.
- Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are co-amplified. *Proc Natl Acad Sci USA* 105: 13486–13491.
- Xiao JH, Wang NX, Li YW, Murphy RW, Wan DG, et al. (2010) Molecular Approaches to identify cryptic species and polymorphic species within a complex community of fig wasps. *PLoS One* 5: e15067.
- Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biol* 2: e312.
- Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol* 16: 314–321.