MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2017) **17**, 796–805 doi: 10.1111/1755-0998.12601

DNA barcodes of the native ray-finned fishes in Taiwan

CHIA-HAO CHANG,* KWANG-TSAO SHAO,* HAN-YANG LIN,* YUNG-CHIEH CHIU,† MAO-YING LEE,* SHIH-HUI LIU‡ and PAI-LEI LIN*

**Biodiversity Research Center, Academia Sinica, 128 Academia Road Sec. 2, Nankang Taipei 115, Taiwan, †Department of Aquaculture, National Taiwan Ocean University, Keelung, Taiwan, ‡Department of Biology, Saint Louis University, 3507 Laclede Avenue, Macelwane Hall, St. Louis MO 63103, USA*

Abstract

Species identification based on the DNA sequence of a fragment of the cytochrome c oxidase subunit I gene in the mitochondrial genome, DNA barcoding, is widely applied to assist in sustainable exploitation of fish resources and the protection of fish biodiversity. The aim of this study was to establish a reliable barcoding reference database of the native ray-finned fishes in Taiwan. A total of 2993 individuals, belonging to 1245 species within 637 genera, 184 families and 29 orders of ray-finned fishes and representing approximately 40% of the recorded ray-finned fishes in Taiwan, were PCR amplified at the barcode region and bidirectionally sequenced. The mean length of the 2993 barcodes is 549 bp. Mean congeneric K2P distance (15.24%) is approximately 10-fold higher than the mean conspecific one (1.51%), but approximately 1.4-fold less than the mean genetic distance between families (20.80%). The Barcode Index Number (BIN) discordance report shows that 2993 specimens represent 1275 BINs and, among them, 86 BINs are singletons, 570 BINs are taxonomically concordant, and the other 619 BINs are taxonomically discordant. Barcode gap analysis also revealed that more than 90% of the collected fishes in this study can be discriminated by DNA barcoding. Overall, the barcoding reference database established by this study reveals the need for taxonomic revisions and voucher specimen rechecks, in addition to assisting in the management of Taiwan's fish resources and diversity.

Keywords: barcode gap, biodiversity, COI, DNA barcoding, mitochondrial genome

Received 5 July 2016; revision received 19 September 2016; accepted 21 September 2016

Introduction

Comprising more than 33 000 species (Eschmeyer 2016), fishes constitute a vast proportion of vertebrate diversity and also play a crucial role in the human food supply (Tacon & Metian 2013). However, fishes are very vulnerable to human activities (Clausen & York 2008). Although approximately 300 new fish species have been identified every year in the past two decades (Eschmeyer 2016), anthropogenic impacts, such as global warming, water extraction, invasive species, overfishing and habitat degradation, have resulted in a devastating loss of fish diversity (Olden *et al.* 2007; Rahel *et al.* 2008; Xenopoulos *et al.* 2005). In view of the value of taxonomy to conserving biodiversity, academics have appealed for more scientists to devote themselves to taxonomical studies (Mora 2014; Mora *et al.* 2008; Reid *et al.* 2013).

Taxonomists have traditionally utilized morphological characters as taxonomic tools to identify fish species. However, during fish development, morphological characters are not always stable and these characters often cannot be assessed in sectioned specimens. Even when experienced taxonomists have intact adult specimens to work with, fish identification may not be straightforward if morphological characters are too subtle or if the existing literature and taxonomic history are contradictory. Lack of taxonomic rigour has impeded sustainable use and conservation of worldwide fish resources (Fischer 2013; Ward *et al.* 2009), so a reliable and efficient means to authenticate fish species is urgently needed.

Molecular identification, that is employing molecular markers to authenticate species, is nowadays widely applied. Initially, this approach employed protein markers (Avise 1975), but now mainly relies on mitochondrial DNA (mtDNA) (Avise 1994). Molecular identification based on mtDNA has several advantages over a morphological approach. First, DNA is more resistant to degradation than morphological characters. For example, DNA extraction is still workable from samples that have undergone food processing or digestion (Chang *et al.* 2013, 2016, 2014; Galal-Khallaf *et al.* 2016; Long *et al.*

Correspondence: Kwang-Tsao Shao, Fax: +886-2-27883463; E-mail: zoskt@gate.sinica.edu.tw

2013; Moran *et al.* 2015; Yang *et al.* 2012). Second, DNA can be extracted from a tiny piece of tissue, including muscle, fin and teeth, so whole intact specimens are not required for identification (Galal-Khallaf *et al.* 2014; Kane & Hellberg 2016; Lee *et al.* 2013; Wen *et al.* 2015; Zhao *et al.* 2013). Third, unlike morphological characters that vary or are absent through distinct developmental stages, resulting in species misidentification (Becker *et al.* 2015; Ko *et al.* 2013; Leis 2015; Lin *et al.* 2016), DNA characters are constant throughout development. Thus, genetic characters can be applied to authenticate fish eggs, larvae, juveniles and adults (Burghart *et al.* 2014; Lee & Kim 2014). Finally, advancements in technology make it quite easy to replicate and determine DNA sequences, while computer software can automatically read data, assess the characters, and compare the resulting sequences so that the training required for a molecular authentication approach is much less that of a morphological one. Molecular identification is being increasingly utilized to tackle many issues including illegal species exploitation, food fraud and biological invasions, and for biodiversity monitoring (Bohmann *et al.* 2014; Collins *et al.* 2013; Frantine-Silva *et al.* 2015; Gonçalves *et al.* 2015; Hubert *et al.* 2015; Khaksar *et al.* 2015; Xiong *et al.* 2016).

Many mitochondrial genes, such as cytochrome *b* (cyt *b*), 16S ribosomal RNA (rRNA) and 12S rRNA, have been utilized as genetic markers for molecular identification (Huang *et al.* 2012; Jogayya *et al.* 2013; Wang *et al.* 2015). However, the accuracy of molecular identification relies on having a reliable and complete reference database (Ratnasingham & Hebert 2007), so inconsistent genetic marker usage could impede the application of molecular authentication. Since Hebert *et al.* (2003) first employed the cytochrome *c* oxidase subunit I (COI, which encodes part of the terminal enzyme of the mitochondrial respiratory chain), for species identification, it has been demonstrated that this genetic fragment could serve as a 'DNA barcode' for biological authentication in many different kinds of animals, from invertebrates to vertebrates (Barrett & Hebert 2005; Clare *et al.* 2007; Hendrich *et al.* 2015; Paknia *et al.* 2015; Ward *et al.* 2005; Waugh 2007). The Fish Barcode of Life Initiative (FISH-BOL) –an internationally coordinated project to establish standardized DNA barcodes for authoritatively verified voucher specimens – was established to construct a reference library for all fishes (Ward *et al.* 2009). Many fish barcoding studies devoted to FISH-BOL from areas across the globe have been conducted (Cawthorn *et al.* 2011; Chakraborty & Ghosh 2014; Karim *et al.* 2016; Keskİn & Atar 2013; Knebelsberger *et al.* 2014, 2015; McCusker *et al.* 2013; Ribeiro *et al.* 2012; Rosso *et al.* 2012; Wang *et al.* 2012; Ward *et al.* 2005; Zhang & Hanner 2012), and almost one-third (11 227 species) of all described fishes have now been barcoded (FISH-BOL 2016). Still, there is a long way to go to complete this project and further barcoding is critical to improving its taxonomic resolution.

Taiwan is a relatively small island but has a rich fish biodiversity. According to the Fish Database of Taiwan (http://fishdb.sinica.edu.tw), the total number of fish species in Taiwan exceeds 3000, so just over 9% of the world's fish species have been recorded in Taiwan. This high fish diversity arises for two main reasons (Shao 2009): (i) Taiwan is located at the northern border of the East Indies representing the world's prime hot spot for marine biodiversity, and also at the apex of the 'The Coral Triangle', so that fish eggs, larva, juveniles and adults are easily transported to Taiwanese waters via the Kuroshio and South China Sea ocean currents; and (ii) Taiwan has various kinds of marine habitat, including mangrove forests, estuaries, sandy barrier lagoons, coral reefs and water depths ranging from relatively shallow to almost 6000 m. Three main ocean currents – Kuroshio, China Coast and South China Sea – intersect in the waters around Taiwan, with water temperatures differing between northern and southern Taiwan by at least six to seven degrees Celsius. Despite its high fish biodiversity, the Taiwanese piscifauna has not been comprehensively barcoded and nor has Taiwan established its own DNA barcode database. The goal of this study was to create a barcode library for Taiwanese fish species and to submit reference barcoding sequences to FISH-BOL in order to promote further research in taxonomy, forensics and ichthyoplankton.

Materials and methods

Sampling

Between 2004 and 2016, a total of 2993 fish specimens was collected. Fishes were identified to species level based on morphological characters by experts and taxonomists, who mainly followed the identification keys of Shen *et al.* (1993) and Nakabo (2013). All of the collected fish species are native to Taiwan, and most of them were gathered from Taiwan and its adjacent waters. As marine fishes usually have wide geographic distributions, some specimens were sampled from Philippine or New Guinean waters. The distribution of all sampling localities is shown in Fig. 1. Two pieces of muscle tissue or fin tissue were removed from each fish specimen: one tissue sample was preserved in 95% ethanol at 4 °C and the other one was kept in liquid nitrogen. The voucher specimens were fixed by formalin and then transferred into 70% ethanol. In some extreme cases, fishes were directly preserved in 95% ethanol if they were very small or, if the fish specimen was too big to retain a biological specimen, a digital image or e-voucher was captured following the

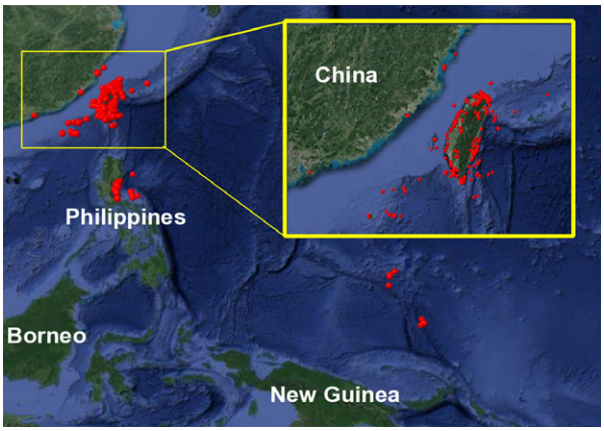

Fig. 1 Distribution of the sampling localities for collected specimens in this study.

FISH-BOL protocol (Steinke & Hanner 2011). All collected specimens, including vouchers and tissue samples, have been deposited at the Biodiversity Research Museum, Academia Sinica, Taiwan, each with its own unique accession number detailing sampling date, place and collector, so that the information is easily accessible on the Fish Database of Taiwan (http://fishdb.sinica.edu.tw). As an example, detailed information for specimen ASIZP0805484 in the Fish Database of Taiwan is shown in Fig. 2.

DNA extraction, PCR and DNA sequencing

Over the course of the 12 years, this project has been ongoing, and DNA extraction, polymerase chain reaction (PCR) and DNA sequence determination protocols have changed as more convenient DNA extraction or PCR kits became available. Whichever kits were utilized, the same four primers (two forward and two reverse) to amplify the DNA barcoding region were consistently employed throughout the entire project. The latest methodology (after 2014) is described as follows: DNA was extracted from each tissue sample using the Quick Gene DNA Tissue Kit S (Fujifilm, Tokyo, Japan). PCR amplifications of the 5′ region of the COI gene (approximately 650 bp) were performed in a mixture with a final volume of 25 μL containing 10–100 ng template DNA, 5 μmol of each forward and reverse primer, forward: FishF1 (5′- T CA ACC AAC CAC AAA GAC ATT GGC AC-3′) and FsihF2 (5′-TCG ACT AAT CAT AAA GAT ATC GGC AC-3′); reverse: FishR1 (5′- TAG ACT TCT GGG TGG CCA AAG AAT CA-3′) and FishR2 (5′-ACT TCA GGG TGA CCG AAG AAT CAG AA-3′) (Ward *et al.* 2005), and uses 12.5 μL of Fast-Run™ Advanced Taq Master Mix (ProTech, Taipei, Taiwan) and distilled water. Thermal cycling began with one cycle at 94 °C for 4 min, followed by 35 cycles of denaturation at 94 °C for 30 s, 45–55 °C (to effect the best balance between PCR productivity and specificity) for 30 s, 72 °C for 30 s and, finally,

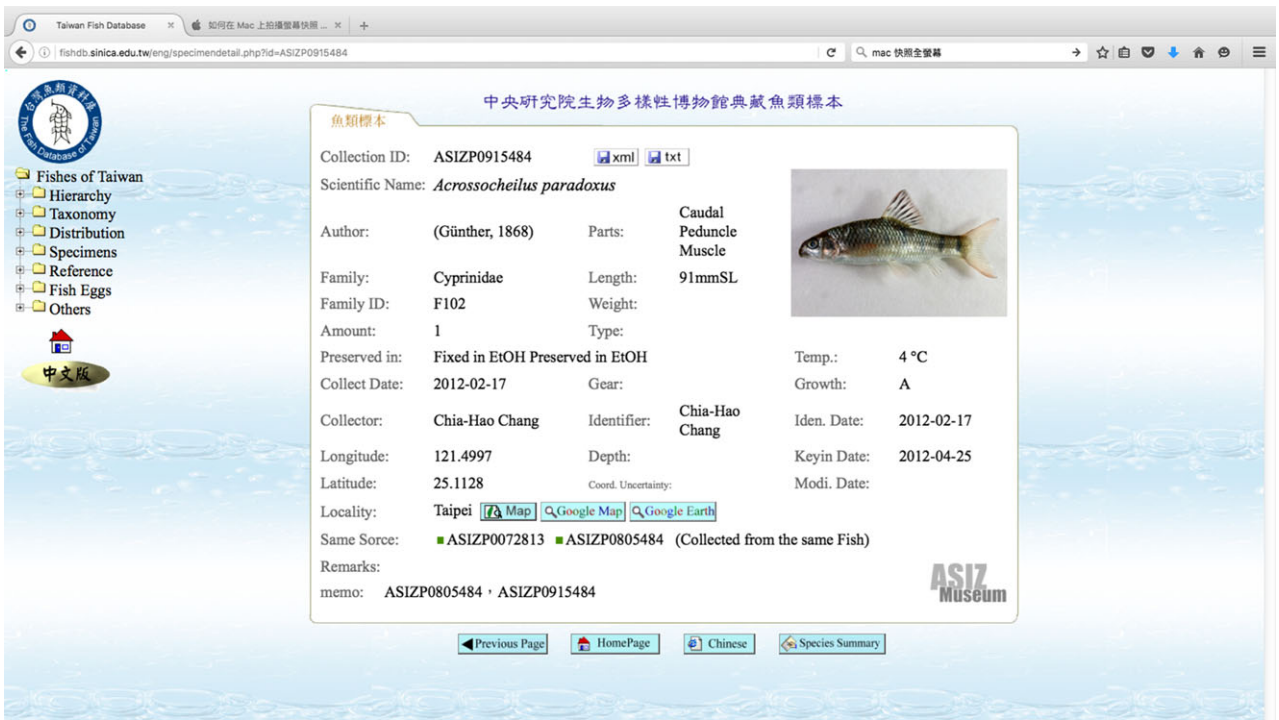

Fig. 2 Specimen data for ASIZP0915484 from the Fish Database of Taiwan. [Colour figure can be viewed at wileyonlinelibrary.com]

a single extension step at 72 °C for 7 min. PCR products were purified using a PCR DNA Fragment Extraction Kit (Geneaid, Taipei, Taiwan). Approximately 50 ng of the purified PCR product was employed as template for sequencing, which we performed following the protocol of the ABI PRISM BIGDYE terminator v3.1 Cycle Sequencing Kit (PE Applied Biosystems, Foster City, CA, USA) with the primers used for PCR (by Mission Biotech Inc., Taipei, Taiwan). The beginnings and ends of the contiguous sequences from both directions of the COI gene from each sample were trimmed, and then, we constructed the contig sequence using the program BIOEDIT ver. 7.1.9. After trimming, all contig sequences started at codon position one and ended at position three and no stop codons were detected. All obtained barcodes are available both in the Barcode of Life Data system (BOLD) and GenBank, and the details are given in Table S1 (Supporting information).

Molecular data analysis

The entire data set has been uploaded to BOLD under project title 'Native teleost fishes in Taiwan'. BOLD version 3.6 analytical tools were employed to conduct the following analyses.

The distance summary with the parameter setting 'muscle alignment algorithm (Edgar 2004); pairwise deletion (ambiguous base/gap handling)' was utilized to estimate the Kimura 2-parameter (K2P) distances for different taxonomical ranks, including species, genus and family levels. Barcode gap analysis, which constructs the distribution of intraspecific and interspecific genetic distances [nearest neighbour (NN) analysis of each species to establish the boundaries of the barcode locus], was performed with the parameter setting 'K2P; muscle alignment algorithm; pairwise deletion (ambiguous base/gap handling)'. The Barcode Index Number (BIN) discordance report was used to disclose cryptic diversity, to verify the accuracy of species identifications, and to detect cases of interspecific haplotype sharing or low levels of genetic differentiation between species. The Taxon ID tree was employed to construct a Neighbour-Joining (NJ) tree of the entire 2993 sequences with the parameter setting 'K2P distance model; muscle alignment algorithm (Edgar 2004); pairwise deletion (ambiguous base/gap handling)'. Moreover, to reveal the phylogenetic relationships of some fish species with bootstrapping support values, the sequences of these fishes were aligned by ClustalW (codon), and then, the NJ tree of K2P distances was constructed with 100 000 bootstrapping replications using Mega6 (Tamura *et al.* 2013).

In order to verify species identification success, this study followed Decru *et al.* (2016) to apply three criteria to the collected barcoding data set, using SPECIESIDENTIFIER v1.7.8 (Meier *et al.* 2006) with a 95% threshold value: best match (BM), best close match (BCM) and all species barcodes (ASB), as proposed by Meier *et al.* (2006). As there were no conspecific barcoding sequences for those fishes having only one sequence (singletons), they were automatically assigned as 'incorrectly identified' under the BM and BCM criteria.

Results

The 2993 fish specimens we barcoded represent 1245 species from 637 genera, 184 families and 29 orders of ray-finned fishes (Actinopterygii). Based on the checklist of the Fish Database of Taiwan, this collection covers 42.59% (1245/2923) of recorded species, 57.34% (637/1111) of recorded genera, 73.60% (184/250) of recorded families and 87.88% (29/33) of the recorded orders of ray-finned fishes in Taiwan. After editing, all barcode sequences ranged in length from 500 to 552 nucleotides (mean: 548.87; SD: 8.19), and the mean nucleotide frequencies of the entire data set are A (23.72%), T (29.51%), G (17.61%) and C (29.16%).

The Taxon ID tree (Fig. S1, Supporting information) reveals that specimens generally formed phylogenetic clusters that reflected prior taxonomic assignment based on morphology. Moreover, the barcode gap analysis shows that 125 species lack a barcode gap (intraspecific K2P distance ≥ interspecific one), 118 species are with high K2P intraspecific distance (>2%), and 167 species are with low K2P distance to another species (≤2%), which indicates that the majority of the studied species could be authenticated by the barcode approach. Actually, only 99 species failed to be discriminated because either the K2P distances between each of them and their own nearest neighbour (NN) were <2%, or the distances to their nearest neighbour were less than the maximum intraspecific distance (Table S2, Supporting information). Overall, the mean K2P distance of a species to its NN was 11.49% (SD: 6.42%).

Mean K2P distances within species, within genera, and within families were 1.51%, 15.24% and 20.80%, respectively (Table 1). The largest intraspecific K2P

Table 1 Summary of K2P genetic distances (%) calculated for different taxonomic levels

				K2P genetic distance (%)		
	N	Taxa	Comparisons	Minimum	Maximum	Mean and SD
Within species	2428	680	4357	0.00	25.01	1.51 ± 4.06
Within genus	2050	220	14 955	0.00	34.29	15.24 ± 5.15
Within family	2670	102	89 136	0.00	41.35	20.80 ± 3.14

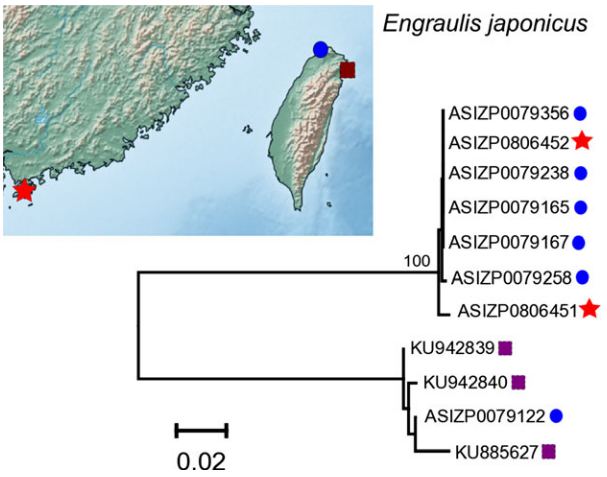

Fig. 3 Neighbour-joining tree of *Engraulis japonicas* from DNA barcode sequences with 100 000 bootstrapping replicates. Taxa are labelled according to BOLD sample IDs and sampling localities: New Taipei, Taiwan (blue circle), Yilan, Taiwan (purple square), and Hong Kong (red star). Bootstrap values >70% are indicated.

distance was present in *Engraulis japonicas* (11 specimens) (Fig. 3, Table S3, Supporting information). For several species, such as *Acanthogobius hasta* (three specimens), *Aphyocypris moltrechti* (five specimens) and *Onychostoma alticorpus* (6 specimens), all specimens carried the same haplotype (Table S3, Supporting information). The mean congeneric distance is approximately 10-fold higher than the mean conspecific one, but approximately 1.4-fold less than the mean genetic distance between families, so mean genetic distance increases with taxonomic level. The BIN discordance report shows that 2993 specimens represent 1275 BINs and, among them, 86 BINs are singletons, 570 BINs are taxonomically concordant, and the other 619 BINs are taxonomically discordant.

For the BM, BCM and ASB analysis of the 2993 sequence data set, which includes the singletons, the proportions of correct identification are 74.64%, 74.64% and 57.66%, respectively; those of ambiguous identifications are 4.94%, 4.94% and 35.31%, respectively; and, finally, those of incorrect identification are 20.41%, 20.41% and 7.01%, respectively. Moreover, for the same three analyses of the data set excluding singletons (2907 sequences), the proportions of correct identification are 76.43%, 76.43% and 59.09%, respectively; those of ambiguous identifications are 4.50%, 4.50% and 34.02%, respectively; and, finally, those of incorrect identification are 19.05%, 19.05% and 6.87%, respectively (Table 2).

Discussion

Lane (2009) proposed that because of the propensity for synonymous nucleotide mutations in the COI, a species rapidly gains a haplotype (or a tight cluster of closely related ones) that is distinctive from those of other congeneric species. The gap between COI intraspecific diversity and interspecific diversity is termed the 'barcode gap', which is crucial for the discriminatory power of DNA barcoding (Ward & Holmes 2007). The barcode gap can be seen in this (Table 1) as well as many other previous fish barcoding studies (KeskİN & Atar 2013; Knebelsberger *et al.* 2014; Pereira *et al.* 2013; Rasmussen *et al.* 2009), confirming yet again that this approach is an efficient way to tell fish species apart. In this study, we have established a DNA barcode reference database, including more than 40% of the recorded native ray-finned fishes in Taiwan. Barcode sequence analyses of our data set reveal that the mean NN distance is approximately 7.5-fold higher than the mean intraspecific one. Most species are arranged into monophyletic units in the NJ tree, again indicating that our barcode database is suitable for discriminating native Taiwanese fishes. The mean intraspecific K2P distance of Taiwanese fishes is higher than that of fish studies from other geographic areas (Cawthorn *et al.* 2011; Dahruddin *et al.* 2016; Karim *et al.* 2016; KeskİN & Atar 2013; Knebelsberger *et al.* 2014, 2015; Ribeiro *et al.* 2012; Rosso *et al.* 2012; Wang *et al.*

Table 2 Results of identification success analysis for the criteria: best match, best close match and all species barcodes (Meier *et al.* 2006)

	Best match (%)	Best close match (%)	All species barcodes (%)
With singletons			
Correct identifications	2234 (74.64%)	2234 (74.64%)	1726 (57.66%)
Ambiguous identifications	148 (4.94%)	148 (4.94%)	1057 (35.31%)
Incorrect identifications	611 (20.41%)	611 (20.41%)	210 (7.01%)
Sequences without any match closer than threshold	NA	NA	NA
Without singletons			
Correct identifications	2222 (76.43%)	2222 (76.43%)	1718 (59.09%)
Ambiguous identifications	131 (4.50%)	131 (4.50%)	989 (34.02%)
Incorrect identifications	554 (19.05%)	554 (19.05%)	200 (6.87%)
Sequences without any match closer than threshold	NA	NA	NA

2012), but it is similar to that of Indian freshwater fishes (Chakraborty & Ghosh 2014). However, our value could be an overestimate or underestimate for the following reasons.

First, broader utilization of genetic technology has shown that many marine fish taxa may be comprised of distinct lineages, indicative of cryptic diversity (Bass *et al.* 2005; Craig *et al.* 2009; Puckridge *et al.* 2013; Tunnicliffe *et al.* 2010; Zemlak *et al.* 2009). In this study, the mean K2P distance within Japanese anchovy (*Engraulis japonicus*) is 12.22% (Table S3, Supporting information), consistent with previous research showing high population variation (Yu *et al.* 2005). Our NJ tree exhibits two monophyletic groups for this species (Fig. 3 and S1, Supporting information), so these specimens warrant careful taxonomic re-examination. Similarly, some species such as *Bleekeria mitsukurii* (mean intraspecific distance, MID = 8.65%), *Callionymus planus* (MID = 8.53%), *Secutor ruconius* (MID = 5.12%) and *Decapterus maruadsi* (MID = 4.86%) have high intraspecific distance values and exhibit polyphyletic groups in the phylogenetic analysis, also suggestive of cryptic diversity.

Second, successful barcoding relies on high-quality DNA sequences and correct scientific naming, so morphological misidentifications of voucher specimens, DNA contamination, mislabelling during specimen processing and incomplete knowledge of the taxonomic literature can contribute to ambiguous barcoding results (Radulovici *et al.* 2010; Tautz *et al.* 2003). Compared to other DNA barcoding studies (Decru *et al.* 2016; Pereira *et al.* 2013), the low identification success rate of our study, even when singletons are removed from the analysed data set, indicates that a re-inspection of the collected fish specimens is warranted and that the ichthyofauna of Taiwan may not yet be comprehensively documented (Table 2). In addition, the BIN discordance report shows that there are 615 taxonomically discordant BINs, and the intraspecific distance values of certain species are larger than their interspecific values (e.g. *Ceratoscopelus warmingii* (Fig. 4), *Cephalopholis miniata*, *Cheilodipterus quinquelineatus*, *Cynoglossus kopsii*) or share haplotypes with other species (e.g. *Kumococius rodericensis*, *Grammoplites scaber*, *Pseudorhombus elevatus*, *Pseudorhombus arsius*, *Trichiurus lepturus*, *Trichiurus japonicus*) (Fig. 5 and S1, Table S2, Supporting information). These specimens require verification, which fortunately is facilitated by their preservation at the Biodiversity Research Museum, Academia Sinica. *Trichiurus lepturus* also has been shown to contain several divergent lineages, which may represent different species (Chakraborty *et al.* 2006; Chakraborty & Iwatsuki 2006; Hsu *et al.* 2007). Thus, our collection of *Trichiurus* specimens, combined with those of other studies, provides researchers with a good resource for a taxonomic review of *T. lepturus*.

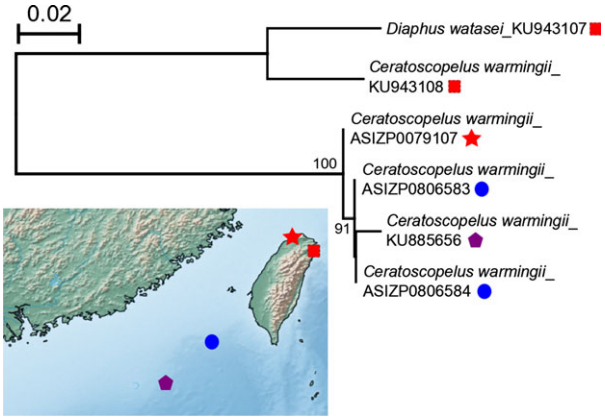

Fig. 4 Neighbour-joining tree of *Ceratoscopelus warmingii* and *Diaphus watasei* from DNA barcode sequences with 100 000 bootstrapping replicates. Taxa are labelled by their scientific names, BOLD sample IDs and sampling localities: New Taipei, Taiwan (red star), Yilan, Taiwan (red square), South China Sea (purple pentagon) and offshore southwestern Taiwan (blue circle). Bootstrap values >70% are indicated.

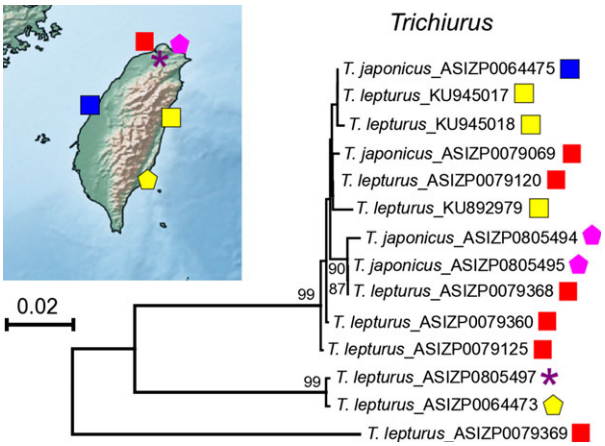

Fig. 5 Neighbour-joining tree of *Trichiurus japonicus* and *T. lepturus* from DNA barcode sequences with 100 000 bootstrapping replicates. Taxa are labelled by their scientific names, BOLD sample IDs and sampling localities: New Taipei, Taiwan (red star), Taipei, Taiwan (purple asterisk), Keelung, Taiwan (pink pentagon), Hualien, Taiwan (yellow square), Chenggong, Taiwan (yellow pentagon), and Changhua, Taiwan (blue square). Bootstrap values >70% are indicated.

Third, failure of DNA barcodes to identify species may be due to incomplete lineage sorting attributable to recent speciation and haplotype sharing through hybridization. It has been reported that some tuna species (genus *Thunnus*) cannot be discriminated by DNA barcoding (Cawthorn *et al.* 2011; Ward *et al.* 2009) as they are genetically similar at the DNA barcode region due to recent and rapid speciation (Viñas & Tudela 2009). Hence, a faster evolving DNA fragment, such as the

mitochondrial control region (mt CR) or ribosomal DNA first internal transcribed spacer (ITS1), may be better for authenticating *Thunnus* fishes (Pedrosa-Gerasmio *et al.* 2012; Viñas & Tudela 2009). In this study, DNA barcodes of specimens of three *Thunnus* fishes (*T. albacares*, *T. alalunga* and *T. orientalis*) were sequenced and the BIN discordance report illustrates that these three species cannot be distinguished (Fig. S1, Supporting information). Thus, a reference database of CR or ITS1 for *Thunnus* fishes is necessary for their molecular authentication. Similarly, hybridization between different damselfishes and *Sillago* fishes has been observed (Coleman *et al.* 2014; Krück *et al.* 2013). Here, our study shows that DNA barcoding failed to identify three *Sillago* fishes (*S. japonica*, *S. sihama* and *S. asiatica*) and two pairs of *Abudefduf* fishes (*A. bengalensis* and *A. septemfasciatus*; *A. vaigiensis* and *A. sexfasciatus*) (Fig. 6 and S1, Supporting information), so these specimens require careful re-examination and nuclear genes should be sequenced in order to establish whether hybridization has occurred. Mullen *et al.* (2012) demonstrated that habitat degradation promotes hybridization between damselfishes, so the ongoing deterioration of coral reefs in Taiwan may not only threaten population size, but also the genetic integrity of Taiwan's damselfishes (Dai *et al.* 2002; Kuo *et al.* 2012; Liu *et al.* 2012). The damselfish collection of this study could be a basis for long-term monitoring of damselfish hybridization.

Global fish diversity is currently highly threatened. Exotic fish species introduced by the aquarium or aquaculture trade has resulted in a worldwide homogenization of ichthyofauna (Blanchet *et al.* 2010; Leprieur *et al.* 2008). Moreover, overexploitation of fishery resources has prompted some scientists to pessimistically predict that all fisheries will have collapsed by 2048 (Worm *et al.* 2006). Species is the unit of biodiversity, so DNA barcoding is widely applied to evaluate fish biodiversity, to monitor fish conservation and to manage fishery resources (Ardura *et al.* 2010; Lewis *et al.* 2016; Takahara *et al.* 2013; Thomsen *et al.* 2012; Valdez-Moreno *et al.* 2012; Weigt *et al.* 2012). Our DNA barcoding study of the native ray-finned fishes in Taiwan is not only beneficial for fish conservation in Taiwan, but also beyond. Knowing where introduced fishes come from assists in preventing further invasions (Lee 2002). A Taiwanese fish, the royal damselfish (*Neopomacentrus cyanomos*), is a newly recorded alien species in the Gulf of Mexico (Robertson *et al.* 2016), and our barcoding data on Taiwanese royal damselfish specimens will be useful in inferring the source of this invasive population and what is the invasion process in the Atlantic.

Undoubtedly, fish diversity in Taiwan has been greatly reduced in recent years. Of the 265 species of native freshwater fishes in Taiwan, nearly 20% are endangered by invasive species or habitat degradation (Chen *et al.* 2012); Chen *et al.* (2015) reported that the marine fish assemblages at two nuclear power plants in northern Taiwan had been remarkably reduced from 100–120 species to 20–30 species in the past 30 years. In addition, a paucity of fishes in reefs was also noticed by the XL Catlin Seaview Survey team in their 2016 Taiwan expedition (http://catlinseaviewsurvey.com/news/16-05-2016/taiwan-thats-a-wrap). Traditionally, monitoring fish diversity consumes a lot of time, money and labour. However, with the ever-expanding barcode database and the growth in biotechnology, such as next-generation sequencing technology and analysis of environmental DNA (eDNA) extracted from water, assessing fish diversity is becoming ever-more efficient (Collins *et al.*

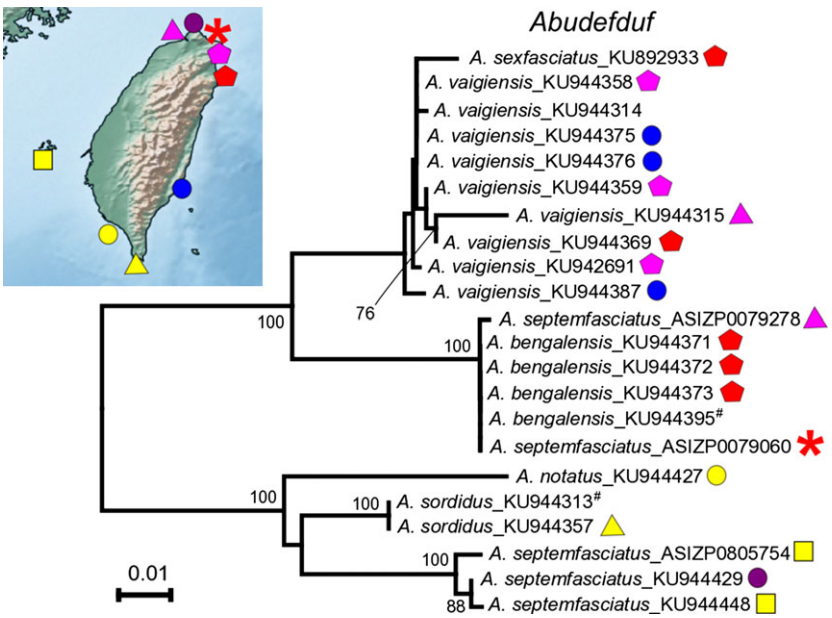

Fig. 6 Neighbour-joining tree of six *Abudefduf* species from DNA barcode sequences with 100 000 bootstrapping replicates. Taxa are labelled by their scientific names, BOLD sample IDs, and sampling localities: New Taipei, Taiwan (pink triangle), Yehliu, Taiwan (purple circle), Keelung, Taiwan (red asterisk), Daxi, Taiwan (pink pentagon), Nanfengao, Taiwan (red pentagon), Changhua, Taiwan (blue circle), Kending, Taiwan (yellow triangle), Liuqiu, Taiwan (yellow circle), and Penghu, Taiwan (yellow square). Bootstrap values >70% are indicated. #Sampling locality is unknown.

2013; Takahara *et al.* 2013; Thomsen *et al.* 2012). As our barcode database of Taiwanese fishes develops, it will make information on Taiwan's fish diversity more accessible than ever. Recently, some fish conservation policies in Taiwan are being strongly debated, including prohibition of the whitebait fishery, a ban on gillnet use in Taiwanese coastal waters and establishment of marine protected areas. All these contentions can be eased somewhat if stakeholders can precisely identify the fish species in their catches or waters, thereby facilitating more sustainable exploitation of fish resources and better protection of fish diversity.

Acknowledgements

This study was supported by several Taiwanese research grants from the Forestry Bureau, Council of Agriculture and Executive Yuan to the corresponding author (93農科-2.4.1-務-e2, 94農科-9.1.7-務-e2(1), 95農科-11.1.3-務-e2(2), 96農科-11.1.2-務-e2(1), 97農科-11.2.1-漁-F1(6), 98-林發-09.1-保-19, 99林管-02.1-保-34, 100農科-8.1.5-務-e1, 101農科-13.5.1-務-e1, 102農科-13.5.1-務-e1, 103林發-08.1-保-23, 104農科-13.8.1-務-e5, and 105農科-12.6.3-務-e1), and by grants from the Ministry of Science and Technology, Taiwan, to the first author (MOST 104-2811-B-001-107 and 105-2917-I-564-075). The authors thank all the students, assistants, and postdoc fellows who were once affiliated to the Laboratory of Fish Ecology and Evolution, and participated in this barcoding project. We are also grateful to Dr. John O'Brien for editing assistance.

Conflict of interests

The authors report no conflict of interests.

References

Ardura A, Linde AR, Moreira JC, Garcia-Vazquez E (2010) DNA barcoding for conservation and management of Amazonian commercial fish. *Conservation Biology*, **143**, 1438–1443.

Avise JC (1975) Systematic value of electrophoretic data. *Systematic Biology*, **23**, 465–481.

Avise JC (1994) *Molecular Markers, Nature History and Evolution*. Chapman & Hall, New York.

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. *Canadian Journal of Zoology*, **83**, 481–491.

Bass AL, Dewar H, Thys T, Streelman JT, Karl SA (2005) Evolutionary divergence among lineages of the ocean sunfish family, Molidae (Tetraodontiformes). *Marine Biology*, **148**, 405–414.

Becker BA, Sales NG, Santos GM, Santos GB, Garvalho DC (2015) DNA barcoding and morphological identification of neotropical ichthyoplankton from the Upper Paraná and São Francisco. *Journal of Fish Biology*, **87**, 159–168.

Blanchet S, Grenouillet G, Beauchard O, *et al.* (2010) Non-native species disrupt the worldwide patterns of freshwater fish body size: implications for Bergmann's rule. *Ecology Letters*, **13**, 421–431.

Bohmann K, Evans A, Gilbert MTP, *et al.* (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, **29**, 358–367.

Burghart SE, Woudenberg LV, Daniels CA, *et al.* (2014) Disparity between planktonic fish egg and larval communities as indicated by DNA barcoding. *Marine Ecology Progress Series*, **503**, 195–204.

Cawthorn D-M, Steinman HA, Witthuhn RC (2011) Establishment of a mitochondrial DNA sequence database for the identification of fish species commercially available in South Africa. *Molecular Ecology Resources*, **11**, 979–991.

Chakraborty M, Ghosh SK (2014) An assessment of the DNA barcodes of Indian freshwater fishes. *Gene*, **537**, 20–28.

Chakraborty A, Iwatsuki Y (2006) Genetic variation at the mitochondrial 16S rRNA gene among *Trichiurus lepturus* (Teleostei: Trichiuridae) from various localities: preliminary evidence of a new Species from West coast of Africa. *Hydrobiologia*, **563**, 501–513.

Chakraborty A, Aranishi F, Iwatsuki Y (2006) Genetic differentiation of *Trichiurus japonicus* and *T. lepturus* (Perciformes: Trichiuridae) based on mitochondrial DNA analysis. *Zoological Studies*, **45**, 419–427.

Chang C-H, Jang-Liaw N-H, Lin Y-S, Fang Y-C, Shao K-T (2013) Authentication of the dried seahorses in traditional Chinese medicine market in Taiwan by molecular forensics. *Journal of Food and Drug Analysis*, **21**, 310–316.

Chang C-H, Yao C-J, Yu H-Y, *et al.* (2014) A molecular forensic method for identifying species composition of processed marine mammal meats. *Journal of Forensic and Legal Medicine*, **23**, 65–69.

Chang C-H, Lin H-Y, Ren Q, Lin Y-S, Shao K-T (2016) DNA barcode identification of fish products in Taiwan: Government-commissioned authentication cases. *Food Control*, **66**, 38–43.

Chen I-S, Tzeng C-S, Shao K-T (2012) *Red Data Book of Freshwater Fishes in Taiwan*. Forestry Bureau, Council of Agriculture, Executive Yuan, Taipei, Taiwan.

Chen H, Liao Y-C, Chen C-Y, *et al.* (2015) Long-term monitoring dataset of fish assemblages impinged at nuclear power plants in northern Taiwan. *Scientific Data*, **2**, 150071.

Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN (2007) DNA barcoding of neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes*, **7**, 184–190.

Clausen R, York R (2008) Global biodiversity decline of marine and freshwater fish: a cross-national analysis of economic, demographic, and ecological influences. *Social Science Research*, **37**, 1310–1320.

Coleman RR, Gaither MR, Kimokeo B, *et al.* (2014) Large-scale introduction of the Indo-Pacific damselfish *Abudefduf vaigiensis* into Hawai'i promotes genetic swamping of the endemic congener *A. abdominalis*. *Molecular Ecology*, **23**, 5552–5565.

Collins RA, Armstrong KF, Holyoake AJ, Keeling S (2013) Something in the water: biosecurity monitoring of ornamental fish imports using environmental DNA. *Biological Invasions*, **15**, 1209–1215.

Craig MT, Graham RT, Torres RA, *et al.* (2009) How many species of goliath grouper are there? Cryptic genetic divergence in a threatened marine fish and the resurrection of a geopolitical species. *Endangered Species Research*, **7**, 167–174.

Dahruddin H, Hutama A, Busson F, *et al.* (2016) Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic coverage, identification accuracy, cryptic diversity and identification of exotic species. *Molecular Ecology Resources*. (in press) [Epub ahead of print].

Dai C-f, Soong K, Chen CA, *et al.* (2002) The status of coral reefs in Taiwan and the conservation problems. In: *Proceedings of IUCN/WCPA – EA-4 Taipei Conference*, pp. 266–276. Academia Sinica, Taipei, Taiwan.

Decru E, Moelants T, Gelas KD, *et al.* (2016) Taxonomic challenges in freshwater fishes: a mismatch between morphology and DNA barcoding in fish of the north-eastern part of the Congo basin. *Molecular Ecology Resources*, **16**, 342–352.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Eschmeyer WN (ed.) (2016) Catalog of Fishes. California Academy of Sciences. Electronic version accessed March 10, 2016. http://research

archive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp.
Fischer J (2013) Fish identification tools for biodiversity and fisheries assessments. *FAO Fisheries and Aquaculture Technical Paper*, **585**, 1–107.
Frantine-Silva W, Sofia SH, Orsi ML, Almeida FS (2015) DNA barcoding of freshwater ichthyoplankton in the neotropics as a tool for ecological monitoring. *Molecular Ecology Resources*, **15**, 1226–1237.
Galal-Khallaf A, Ardura A, Mohammed-Geba K, Borrell YJ, Garcia-Vazquez E (2014) DNA barcoding reveals a high level of mislabeling in Egyptian fish fillets. *Food Control*, **46**, 441–445.
Galal-Khallaf A, Ardura A, Borrell YJ, Garcia-Vazqueza E (2016) Towards more sustainable surimi? PCR-cloning approach for DNA barcoding reveals the use of species of low tropic level and aquaculture in Asian surimi. *Food Control*, **61**, 62–69.
Gonçalves PFM, Oliveira-Marques AR, Matsumoto TE, Miyaki CY (2015) DNA barcoding identifies illegal parrot trade. *Journal of Heredity*, **106**, 560–564.
Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.
Hendrich L, Morinière J, Haszprunar G, *et al.* (2015) A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology Resources*, **15**, 795–818.
Hsu K-C, Shih N-T, Ni I-H, Shao K-T (2007) Genetic variation in *Trichiurus lepturus* (Perciformes: Trichiuridae) in waters off Taiwan: several species or cohort contribution. *The Raffles Bulletin of Zoology*, **14**, 215–220.
Huang K-M, Liu S-M, Huang Y-W, Chen T-Z (2012) Identification of species in commercial frozen shrimp meat in Taiwan. *Journal of Food and Drug Analysis*, **20**, 839–983.
Hubert N, Espiau B, Meyer C, Planes S (2015) Identifying the ichthyoplankton of a coral reef using DNA barcodes. *Molecular Ecology Resources*, **15**, 57–67.
Jogayya KN, Meganathan PR, Dubey B, Haque I (2013) Mitochondrial 16S ribosomal RNA gene for forensic identification of crocodile species. *Journal of Forensic and Legal Medicine*, **20**, 334–338.
Kane DE, Hellberg RS (2016) Identification of species in ground meat products sold on the U.S. commercial market using DNA-based methods. *Food Control*, **59**, 158–163.
Karim A, Iqbal A, Akhtar R, *et al.* (2016) Barcoding of fresh water fishes from Pakistan. *Mitochondrial DNA Part A*, **27**, 1–4.
KeskİN E, Atar HH (2013) DNA barcoding commercially important fish species of Turkey. *Molecular Ecology Resources*, **13**, 788–797.
Khaksar R, Carlson T, Schaffner DW, *et al.* (2015) Unmasking seafood mislabeling in U.S. markets: DNA barcoding as a unique technology for food authentication and quality control. *Food Control*, **56**, 71–76.
Knebelsberger T, Landi M, Neumann H, *et al.* (2014) A reliable DNA barcode reference library for the identification of the North European shelf fish fauna. *Molecular Ecology Resources*, **14**, 1060–1071.
Knebelsberger T, Dunz AR, Neumann D, Geiger MF (2015) Molecular diversity of Germany's freshwater fishes and lampreys assessed by DNA barcoding. *Molecular Ecology Resources*, **15**, 562–572.
Ko H-L, Wang Y-T, Chiu T-S, *et al.* (2013) Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. *PLoS ONE*, **8**, e53415.
Krück NC, Tibbetts IR, Ward RD, *et al.* (2013) Multi-gene barcoding to discriminate sibling species within a morphologically difficult fish genus (*Sillago*). *Fisheries Research*, **143**, 39–46.
Kuo C-Y, Yuen YS, Meng P-J, *et al.* (2012) Recurrent disturbances and the degradation of hard coral communities in Taiwan. *PoLS One*, **7**, e44364.
Lane N (2009) Biodiversity: on the origin of bar codes. *Nature*, **462**, 272–274.
Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, **17**, 386–391.
Lee SJ, Kim J-K (2014) Identification of *Trichiurus* (Pisces: Trichiuridae) eggs and larvae from Korea, with a taxonomic note. *Fisheries and Aquatic Sciences*, **17**, 137–143.
Lee E-j, Lee Y-h, Moon S-h, *et al.* (2013) The identification of elephant ivory evidences of illegal trade with mitochondrial cytochrome b gene and hypervariable D-loop region. *Journal of Forensic and Legal Medicine*, **20**, 174–178.
Leis JM (2015) Taxonomy and systematics of larval Indo-Pacific fishes: a review of progress since 1981. *Ichthyological Research*, **62**, 9–28.
Leprieur F, Beauchard O, Blanchet S, Oberdorff T, Brosse S (2008) Fish invasions in the world's river systems: when natural processes are blurred by human activities. *PLoS Biology*, **6**, 404–410.
Lewis LA, Richardson DE, Zakharov EV, Hanner R (2016) Integrating DNA barcoding of fish eggs into ichthyoplankton monitoring programs. *Fishery Bulletin*, **114**, 153–165.
Lin H-Y, Chiu M-Y, Shih Y-M, *et al.* (2016) Species composition and assemblages of ichthyoplankton during summer in East China Sea. *Continental Shelf Research*, **126**, 64–78.
Liu P-J, Meng P-J, Liu L-L, Wang J-T, Leu M-Y (2012) Impacts of human activities on coral reef ecosystems of southern Taiwan: a long-term study. *Marine Pollution Bulletin*, **64**, 1129–1135.
Long BL, Kurta A, Clemans DL (2013) Analysis of DNA from feces to identify prey of big brown bats (*Eptesicus fuscus*) caught in Apple Orchards. *The American Midland Naturalist*, **170**, 287–297.
McCusker MR, Denti D, Guelpen LV, Kenchington E, Bentzen P (2013) Barcoding Atlantic Canada's commonly encountered marine fishes. *Molecular Ecology Resources*, **13**, 177–188.
Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.
Mora C (2014) *Ecology of Fishes on Coral Reefs: The Functioning of an Ecosystem in a Changing World*. Cambridge University Press, Cambridge, UK.
Mora C, Tittensor DP, Myers RA (2008) The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. *Proceedings of the Royal Society B: Biological Science*, **275**, 149–155.
Moran Z, Orth DJ, Schmitt JD, Hallerman EM, Aguilar R (2015) Effectiveness of DNA barcoding for identifying piscine prey items in stomach contents of piscivorous catfishes. *Environmental Biology of Fishes*, **99**, 161–167.
Mullen SP, Little K, Draud M, Brozek J, Itzkowitz M (2012) Hybridization among Caribbean damselfish species correlates with habitat degradation. *Journal of Experimental Marine Biology and Ecology*, **416–417**, 221–229.
Nakabo T (2013) *Fishes of Japan with Pictorial Keys to the Species*, 3rd edn. Tokai University Press, Hanado, Japan.
Olden JD, Hogan ZS, Zanden MJV (2007) Small fish, big fish, red fish, blue fish: size-biased extinction risk of the world's freshwater and marine fishes. *Global Ecology and Biogeography*, **16**, 694–701.
Paknia O, Bergmann T, Hadrys H (2015) Some 'ant'swers: application of a layered barcode approach to problems in ant taxonomy. *Molecular Ecology Resources*, **15**, 1262–1274.
Pedrosa-Gerasmio IR, Babaran RP, Santos MD (2012) Discrimination of juvenile yellowfin (*Thunnus albacares*) and bigeye (*T. obesus*) tunas using mitochondrial DNA control region and liver morphology. *PoLS One*, **7**, e35604.
Pereira LH, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genetics*, **14**, 20.
Puckridge M, Andreakis N, Appleyard SA, Ward RD (2013) Cryptic diversity in flathead fishes (Scorpaeniformes: Platycephalidae) across the Indo-West Pacific uncovered by DNA barcoding. *Molecular Ecology Resources*, **13**, 32–42.
Radulovici AE, Archambault P, Dufresne F (2010) DNA barcodes for marine biodiversity: moving fast forward? *Diversity*, **2**, 450–472.
Rahel FJ, Bierwagen B, Taniguchi Y (2008) Managing aquatic species of conservation concern in the face of climate change and invasive species. *Conservation Biology*, **22**, 551–561.
Rasmussen RS, Morrissey MT, Hebert PDN (2009) DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmo*) from North America. *Journal of Agricultural and Food Chemistry*, **57**, 8379–8385.
Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Reid GM, MacBeath TC, Csatádi K (2013) Global challenges in freshwater-fish conservation related to public aquariums and the aquarium industry. *International Zoo Yearbook*, **47**, 6–45.

Ribeiro AdO, Caires RA, Mariguela TC, *et al*. (2012) DNA barcodes identify marine fishes of São Paulo State, Brazil. *Molecular Ecology Resources*, **12**, 1012–1020.

Robertson DR, Simoes N, Rodríguez CG, España HP (2016) An Indo-Pacific damselfish well established in the southern Gulf of Mexico: prospects for a wider, adverse invasion. *Journal of the Ocean Science Foundation*, **19**, 1–17.

Rosso JJ, Mabragaña E, Castro MG, Astarloa JMDd (2012) DNA barcoding Neotropical fishes: recent advances from the Pampa Plain, Argentina. *Molecular Ecology Resources*, **12**, 999–1011.

Shao K-T (2009) Marine biodiversity and fishery sustainability. *Asia Pacific Journal of Clinical Nutrition*, **18**, 527–531.

Shen SC, Lee SC, Shao KT, *et al*. (1993) *Fishes of Taiwan*. Department of Zoology, National Taiwan University, Taipei.

Steinke D, Hanner R (2011) The FISH-BOL collaborators' protocol. *Mitochondrial DNA*, **22**, 10–14.

Tacon AGJ, Metian M (2013) Fish matters: importance of aquatic foods in human nutrition and global food supply. *Reviews in Fisheries Science*, **21**, 22–38.

Takahara T, Minamoto T, Doi H (2013) Using environmental DNA to estimate the distribution of an Invasive fish species in ponds. *PLoS ONE*, **8**, e56584.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolution genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2775–2729.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution*, **18**, 70–74.

Thomsen PF, Kielgast J, Iversen LL, *et al*. (2012) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, **7**, e41732.

Tunnicliffe V, Koop BF, Tyler J, So S (2010) Flatfish at seamount hydrothermal vents show strong genetic divergence between volcanic arcs. *Marine Ecology*, **31**, 158–167.

Valdez-Moreno M, Quintal-Lizama C, Gómez-Lozano R, García-Rivas MdC (2012) Monitoring an alien invasion: DNA barcoding and the identification of lionfish and their prey on coral reefs of the Mexican Caribbean. *PLoS ONE*, **7**, e36636.

Viñas J, Tudela S (2009) A validated methodology for genetic identification of tuna species (genus *Thunnus*). *PLoS ONE*, **4**, e7606.

Wang Z-D, Guo Y-S, Liu X-M, Fan Y-B, Liu C-W (2012) DNA barcoding South China Sea fishes. *Mitochondrial DNA*, **23**, 405–410.

Wang L-P, Geng R-Q, Liu Z-Q (2015) An effective strategy for species identification of avian meats using the mitochondrial 12S rRNA gene fragment. *Mitochondrial DNA*, **26**, 171–174.

Ward RD, Holmes BH (2007) An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (*cox1*) in fishes. *Molecular Ecology Notes*, **7**, 899–907.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **360**, 1847–1857.

Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, **74**, 329–356.

Waugh J (2007) DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*, **29**, 188–197.

Weigt LA, Baldwin CC, Driskell A, *et al*. (2012) Using DNA barcoding to assess Caribbean reef fish biodiversity: expanding taxonomic and geographic coverage. *PLoS ONE*, **7**, e41059.

Wen J, Zeng L, Sun Y, *et al*. (2015) Authentication and traceability of fish maw products from the market using DNA sequencing. *Food Control*, **55**, 185–189.

Worm B, Barbier EB, Beaumont N, *et al*. (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science*, **314**, 787–790.

Xenopoulos MA, Lodge DM, Alcamo J, *et al*. (2005) Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biology*, **11**, 1557–1564.

Xiong X, Guardone L, Giusti A, *et al*. (2016) DNA barcoding reveals chaotic labeling and misrepresentation of cod (鱈, Xue) products sold on the Chinese market. *Food Control*, **60**, 519–532.

Yang Y-C, Huang Y-W, Hsieh C-H, Huang Y-R, Chen C-H (2012) A unique specification method for processed unicorn filefish products using a DNA barcode marker. *Food Control*, **25**, 292–302.

Yu Z-n, Kong X-y, Guo T-h, *et al*. (2005) Mitochondrial DNA sequence variation of Japanese anchovy *Engraulis japonicus* from the Yellow Sea and East China Sea. *Fisheries Science*, **71**, 299–307.

Zemlak TS, Ward RD, Connell AD, Holmes RH, Hebert PDN (2009) DNA barcoding reveals overlooked marine fishes. *Molecular Ecology Resources*, **9**, 237–242.

Zhang J, Hanner R (2012) Molecular approach to the identification of fish in the South China Sea. *PLoS ONE*, **7**, e30621.

Zhao W, Zhao Y, Pan Y, *et al*. (2013) Authentication and traceability of *Nibea albiflora* from surimi products by species-specific polymerase chain reaction. *Food Control*, **31**, 97–101.

Author Contributions K.-T. Shao designed the study, and C.-H. Chang wrote the article, and C.-H. Chang, H.-Y. Lin, Y.-C. Chiu, M.-Y. Lee and P.-L. Lin collected the fish specimens and determined the barcoding sequences, and C.-H. Chang, H.-Y. Lin, Y.-C. Chiu, M.-Y. Lee, and S.-H. Liu performed computational analyses.

Data accessibility

All collecting and sequence data are available on the Barcode of Life Datasystem (BOLD) under the project 'Native teleost fishes in Taiwan'. DNA sequences were uploaded to NCBI with the following Accession nos.: KU885581–KU885680, KU892792–KU893088, KU942678–KU945272 and KX421780 (see Table S1, Supporting information for the museum ID, BOLD sample ID, BOLD process ID and NCBI accession numbers of each specimen). Alignment of all 2993 sequences and the taxon ID tree (neighbour-joining, K2P) in Newick format have both been uploaded to DRYAD (doi:10.5061/dryad.n36st).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Taxon ID tree of the entire 2993 sequences. Each taxon is labelled with its scientific name, BOLD sample ID, and BIN URIs.

Table S1 Species, museum ID, BOLD sample ID, BOLD process ID, and GenBank accession number of all specimens analysed in this study.

Table S2 Mean and maximum intraspecific values for each specimen compared to the nearest neighbour (NN) K2P distance.

Table S3 Detail of pairwise intraspecific K2P distances for the analysed data set.