

# Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic coverage, identification accuracy, cryptic diversity and identification of exotic species

HADI DAHRUDDIN,\* ADITYA HUTAMA,† FRÉDÉRIC BUSSON,‡ SOPIAN SAURI,\*  
ROBERT HANNER,§ PHILIPPE KEITH,‡ RENNY HADIATY\* and NICOLAS HUBERT¶

\*LIPI, Research Center for Biology, Zoology Division, MZB, Gedung Widiasatwaloka, Jl. Raya Jakarta Bogor Km. 46, Cibinong-Bogor 16911, Indonesia, †Faculty of Mathematics and Natural Science, Animal Bioscience, Bogor Agricultural University, Jl. Raya Darmaga, Bogor 16680, Indonesia, ‡Muséum national d'Histoire naturelle, UMR 7208 (MNHN-CNRS-UPMC-IRD-UCBN), CP 026, 43, rue Cuvier, Paris Cedex 05 F-75231, France, §Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, Guelph, ON, Canada, ¶Institut de Recherche pour le Développement, UMR 226 ISEM (UM2-CNRS-IRD), Place Eugène Bataillon, CC 065, F-34095 Montpellier cedex 05, France

## Abstract

Among the 899 species of freshwater fishes reported from Sundaland biodiversity hotspot, nearly 50% are endemics. The functional integrity of aquatic ecosystems is currently jeopardized by human activities, and landscape conversion led to the decline of fish populations in several part of Sundaland, particularly in Java. The inventory of the Javanese ichthyofauna has been discontinuous, and the taxonomic knowledge is scattered in the literature. This study provides a DNA barcode reference library for the inland fishes of Java and Bali with the aim to streamline the inventory of fishes in this part of Sundaland. Owing to the lack of available checklist for estimating the taxonomic coverage of this study, a checklist was compiled based on online catalogues. A total of 95 sites were visited, and a library including 1046 DNA barcodes for 159 species was assembled. Nearest neighbour distance was 28-fold higher than maximum intraspecific distance on average, and a DNA barcoding gap was observed. The list of species with DNA barcodes displayed large discrepancies with the checklist compiled here as only 36% (i.e. 77 species) and 60% (i.e. 24 species) of the known species were sampled in Java and Bali, respectively. This result was contrasted by a high number of new occurrences and the ceiling of the accumulation curves for both species and genera. These results highlight the poor taxonomic knowledge of this ichthyofauna, and the apparent discrepancy between present and historical occurrence data is to be attributed to species extirpations, synonymy and misidentifications in previous studies.

**Keywords:** DNA barcoding, fish, fisheries management, habitat degradation, wildlife management

Received 15 August 2015; revision received 22 October 2015; accepted 29 October 2015

## Introduction

Amongst the 25 biodiversity hotspots identified worldwide for their remarkable levels of endemism and anthropogenic threats, four are observed in South-East Asia (SEA), including Indo-Burma (Thailand, Cambodia, Laos, Vietnam and Myanmar), Sundaland (Malaysia, Indonesia), Wallacea (Indonesia) and the Philippines (Myers *et al.* 2000). The Sundaland and Indo-Burma hotspots exhibit the highest species richness and endemism in SEA that rank them as some of the world most speciose together with the Brazil's cerrado or West African

forests (Lamoureux *et al.* 2006). These high levels of endemism and species richness may be seen as the result of the complex history of the Indo-Australian Archipelago (IAA) that has been repeatedly fragmented during its ontogenesis through tectonic events (Lohman *et al.* 2011) and eustatic fluctuations (Woodruff 2010). The Sundaland hotspot, however, is currently one of the world's most endangered and aquatic ecosystems exemplify the diversity of anthropogenic pressures faced by its biotas (Hoffman *et al.* 2010). Their functional integrity is currently jeopardized by interactions among ecological (e.g. mining, logging, land conversion, organic and inorganic contaminations) and biotic perturbations (e.g. alien species, overexploitation by inland fisheries), resulting in the destruction of foraging and spawning grounds and

Correspondence: Nicolas Hubert, Fax: 33 4 6714 3622; E-mail: nicolas.hubert@ird.fr

the decline of populations (Schilthuis *et al.* 2005; Clements *et al.* 2006, 2008; Fraser 2006; Normile 2010; Sodhi *et al.* 2010).

Indonesia exhibits one of the world highest densities of freshwater fish species (i.e. 0.6 species per 1000 km<sup>2</sup>) ahead of Brazil (0.37 species per 1000 km<sup>2</sup>) and the Democratic Republic of Congo (0.48 species per 1000 km<sup>2</sup>) (Hubert *et al.* 2015). In Sundaland, nearly 900 species and 430 endemics have been reported, a diversity that accounts for 74% and 48% of the global and endemic diversity, respectively (Hubert *et al.* 2015). Amongst the three major islands of Sundaland, Java exhibits the highest density of species with 1.7 species per 1000 km<sup>2</sup> (ca. 213 species) ahead of Kalimantan (1.2 species per 1000 km<sup>2</sup> for a total of 646 species) and Sumatra (1.7 species per 1000 km<sup>2</sup> for a total of 460 species). During the last decades, the exponential growth of the human population in Java – 130 millions of people sharing 130 000 km<sup>2</sup> – had dramatic consequences on Javanese ecosystems and Javanese biotas are currently some of most threaten in Sundaland.

The diversity loss in Sundaland hotspot is of great concern; however, the knowledge about species taxonomy and distribution is still incomplete for many taxa, what arguably bridles the establishment of sounds conservation plans. This situation is further amplified by the complexity of accurately delineating species in mega-diverse faunas due to the high number of closely related and morphologically similar species (Smith *et al.* 2005; Hubert *et al.* 2012; Jaafar *et al.* 2012), and Indonesian freshwater fishes are no exception (Kadarusman *et al.* 2012). In addition, several recent biogeographic studies have evidenced a substantial of cryptic diversity in Sundaland freshwater fishes frequently displaying intricate patterns of range distribution straddling across islands and calling for regional management plans (Nguyen *et al.* 2008; Pouyaud *et al.* 2009; De Bruyn *et al.* 2013). The

objective of the present study is to provide a DNA barcode reference library of the freshwater fishes of Java and Bali islands with the aim to streamline the ongoing inventory of the Javanese and Balinese ichthyofaunas and to promote more sustainable practices for further taxonomic studies (Hubert & Hanner 2015). The checklist of the Javanese and Balinese freshwater fishes has been assembled from several online catalogues, and implications of this study on the inventory of the Java and Bali ichthyofauna are discussed.

## Materials and methods

### *Specimen collections and identifications*

A total of 3310 specimens, including 162 species, 110 genera and 53 families, have been collected across 95 sites in Java and Bali islands between November 2012 and May 2015 (Fig. 1). Specimens were captured using various gears including electrofishing, seine nets, cast nets and gill nets across sites encompassing the diversity of freshwater lentic and lotic habitats from outlets (i.e. sea level), floodplains, lakes and ponds to upstream tributaries (i.e. 1068 m). Specimens were identified following available monographs (Kottelat *et al.* 1993; Rachmatika 2003; Larson 2009, 2010; Keith *et al.* 2010, 2013), and species names were further validated based on several online catalogues (Froese & Pauly 2011; Eschmeyer & Fricke 2014). Specimens were photographed and individually labelled, and voucher specimens were preserved in a 5% formalin solution. A fin clip or a muscle biopsy was taken for each specimen and fixed in a 96% ethanol solution for further genetic analyses. Both tissues and voucher specimens were deposited in the national collections at the Muzeum Zoologicum Bogoriense (MZB) in the Research Centre for Biology (RCB) from the Indonesian Institute of Sciences (LIPI).

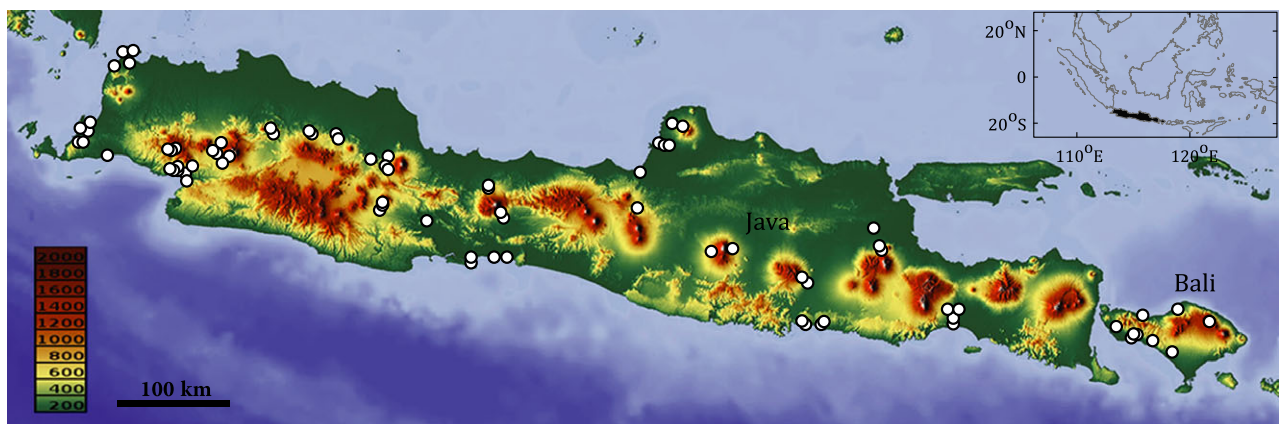


Fig. 1 Collection sites for the 1046 samples analysed in this study. Each point may represent several collection sites.

### *Assembling a checklist of the Javanese and Balinese freshwater fishes*

The checklist of the Javanese and Balinese freshwater fishes has been assembled from available online catalogues including FISHBASE (Froese & Pauly 2011) and Eschmeyer catalogue of fishes (Eschmeyer & Fricke 2014) as detailed in Hubert and colleagues (Hubert *et al.* 2015). Occurrences were further refined based on available monographs (Kottelat *et al.* 1993; Kottelat 2013). This checklist was used to estimate the taxonomic coverage of the present DNA barcoding campaign and flag potential new occurrences. Additional information was included as follows: (i) authors of the original description, (ii) maximum observed length, (iii) type of length measurement, (iv) status of the species namely native or introduced, (v) source references for the distribution, (vi) status of the distribution including endemic, occurring in other countries or original distribution range for introduced species, (vii) occurrence in Java, (viii) potential new occurrence in Java, (ix) occurrence in Bali and (x) potential new occurrence in Bali. The database is provided as an online supplementary material (Table S1, Supporting information).

### *DNA barcode sequencing*

A total of 1403 specimens were selected for sequencing in order to cover as much as possible of (i) the intraspecific genetic diversity by selecting specimens throughout the occurrence range, and (ii) the ontogenetic stages for each species by including juveniles, subadults and adults. Genomic DNA was extracted by a Qiagen DNeasy 96 tissue extraction kit following the manufacturer's specifications. The standard 652-bp segment from the 5' region of the cytochrome oxidase I gene (COI) was amplified using primer cocktail C\_FishF1t1/C\_FishR1t1 including a M13 tails (Ivanova *et al.* 2007). PCR amplifications were performed on a Veriti 96-well Fast thermocycler (ABI – Applied Biosystems) with a final volume of 10.0  $\mu\text{L}$  containing 5.0  $\mu\text{L}$  buffer 2X, 3.3  $\mu\text{L}$  ultrapure water, 1.0  $\mu\text{L}$  each primer (10  $\mu\text{M}$ ), 0.2  $\mu\text{L}$  enzyme Phire<sup>®</sup> Hot Start II DNA polymerase (5U) and 0.5  $\mu\text{L}$  of DNA template (~50 ng). Amplifications were conducted as follows: initial denaturation at 98 °C for 5 min followed by 30 cycles of denaturation at 98 °C for 5 s, annealing at 56 °C for 20 s and extension at 72 °C for 30 s, followed by a final extension step at 72 °C for 5 min. The PCR products were purified with ExoSap-IT<sup>®</sup> (USB Corporation, Cleveland, OH, USA) and sequenced in both directions. Sequencing reactions were performed using the 'BigDye<sup>®</sup> TERMINATOR v3.1 Cycle Sequencing Ready Reaction', and sequencing was performed on the automatic

sequencer ABI 3130 DNA Analyzer (Applied Biosystems).

DNA barcodes, photographs, sequences and collection data were deposited on the Barcode of Life Data System (BOLD) in the projects 'Barcoding Indonesian Fishes – part II. Inland fishes of Java and Bali [BIFB]', 'Barcoding Indonesian Fishes – part III. Sicydiinae of Sundaland [BIFC]', 'Barcoding Indonesian Fishes – part VIb. Widespread primary freshwater fishes of Java and Bali [BIFGA]', 'Barcoding of Indonesian Fishes – part VIIb *Rasbora* spp [BIFHB]' in the container 'Barcoding Indonesian Fishes' of the 'Barcoding Fish (FishBOL)' campaign.

### *DNA barcode analysis*

DNA sequence divergence was calculated using the Kimura 2-parameter (K2P) model (Kimura 1980). The mid-point-rooted neighbour-joining (NJ) tree of K2P distances was constructed to provide a graphic representation of the species divergence as implemented in the Sequence Analysis module of BOLD (Ratnasingham & Hebert 2007). Sequence divergence was considered below and above species boundaries by calculating the maximum intraspecific distance and the distance to the closest phylogenetic neighbour in the data set. The distribution of both distances was examined through sequence divergence class of 1% in order to check for a potential overlap between intraspecific and interspecific sequence divergence. We further checked for a DNA barcoding gap in our data (Meyer & Paulay 2005). Instead of considering potential overlap in the distribution of sequence divergences, we examined the relationships between the maximum intraspecific distance and the distance to the nearest neighbour as it displays potential overlap on an individual basis considering potential overlap for each species (Blagoev *et al.* 2016). Sequence divergence within species, and the potential occurrence of cryptic diversity, was further explored through the Refined Single Linkage (RESL) algorithm to reach decision on the number of operational taxonomic units (OTUs) referenced as Barcode Index Numbers (BIN) in BOLD (Ratnasingham & Hebert 2013).

The taxonomic coverage of the present DNA barcoding campaign was estimated according to the checklist assembled here, and its completeness was further explored through the accumulation curve analysis implemented in the Sequence Analysis module in BOLD (Ratnasingham & Hebert 2007). Accumulation curves were established independently for species and genera across 100 iterations. The taxonomic coverage was further explored through the distribution of the percentage of sampled and nonsampled species across 10-cm size class for euryhaline and amphidromous or primary freshwater families independently.

## Results

### DNA barcode analyses and BIN splits

A total of 1046 sequences belonging to 159 species, 107 genera and 50 families were successfully obtained (Table S2, Supporting information). All the sequences were above 500 bp of length and no codon stops were detected, suggesting that the DNA barcodes collected represent functional coding regions. Although the cocktail of primers failed to amplify 357 specimens (i.e. 25%), they were effective in amplifying and sequencing 159 of the 162 species sampled (i.e. 98%) as no DNA barcodes were recovered for only three species, including *Lethrinus fulviflamma*, *Megalops cyprinoides* and an unidentified South American Loricariid assigned to the genus *Ancistrus*.

Intraspecific distances ranged from 0 to 12.81% and averaged 0.57%, while interspecific distance within genus was 18-fold higher, ranging from 1.56% to 23.7% and averaging 10.21% (Table 1). Although the distribution of the maximum intraspecific distance and the distance to the nearest neighbour overlapped (Fig. 2A,B), nearest neighbour distances were 28-fold higher on average than maximum intraspecific distances (Table S3, Supporting information). Their relationship indicated that maximum intraspecific distances only exceeded nearest neighbour distance in a single species pair and a barcoding gap was generally observed (Fig. 2C). Several species exhibited high maximum intraspecific distances, and BIN counts were higher than species count in 18 cases (Table 2). The highest BIN divergences were observed in *Eleotris fusca* with 12.81% between two BINs and *Giuris margaritacea* with 12.56% between two BINs. The lowest BIN divergences were observed in *Acentrogobius caninus* (two BINS, 1.58%) and *Trichopodus trichopterus* (two BINS, 1.2%). The highest number of BIN was observed in *Channa gachua* with five BINS with a maximum divergence of 5.63% (Table 2).

**Table 1** Summary statistics of the genetic distances (K2P) through increasing taxonomic levels

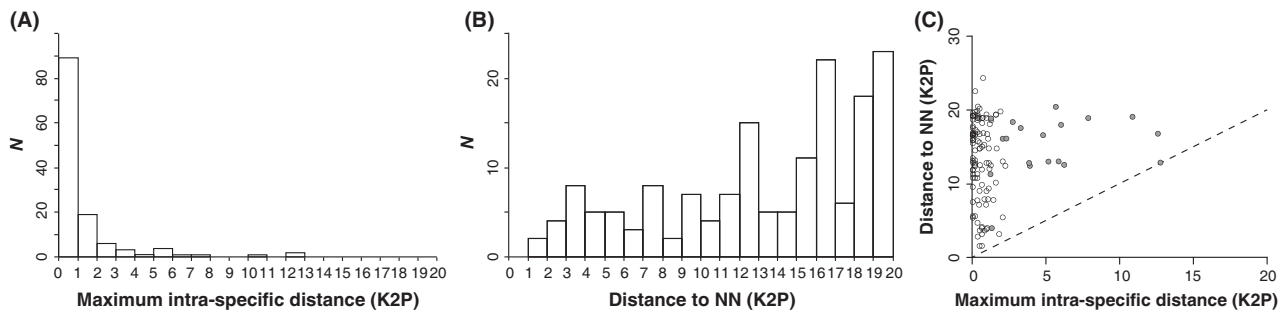
	N (Taxa/ Specimens)	Minimum distance (K2P)	Mean distance (K2P)	Maximum distance (K2P)
Within species	125/1012	0	0.57	12.81
Within genus	36/698	1.56	10.21	23.7
Within family	16/813	6.35	19.35	33.02

### Species diversity and exotic species

The checklist of the freshwater fishes of Java and Bali yielded a total of 227 species belonging to 181 genera and 66 families (Table S1, Supporting information) among which 216 and 40 have been reported from Java and Bali, respectively (Table 3). The species list obtained from the present DNA barcoding campaign, however, poorly matched the checklist compiled from online catalogues. Amongst the 159 species of the reference library, 85 species have been previously reported from Java and Bali, while 74 species correspond to new records for both islands. In Java, DNA barcodes of only 77 species were recovered of the 227 species previously reported (i.e. 36%), while 75 species with DNA barcodes correspond to new records, increasing the species richness of Java by 130%. A similar trend is observed for Bali as amongst the 40 species previously reported, DNA barcodes were recovered for 24 species (i.e. 60%), while 34 species with DNA barcodes correspond to new records, increasing the diversity by 185%.

The accumulation curves displayed an asymptotic trend, indicating that the present sampling was representative of the Java and Bali ichthyofauna, whatever the taxonomic level considered (Fig. 3A). We further examined the distribution of the taxonomic gap in the present library through size class of 10 cm. The Javanese and Balinese ichthyofauna are dominated by species smaller than 20 cm, representing nearly 70% of the overall species pool (Fig. 3B). The distribution of the taxonomic gap through size class varied between euryhaline or amphidromous and primary freshwater families (Fig. 3C,D). A poor taxonomic coverage was recovered at the largest size classes for primary freshwater family (Fig. 3D), as 12 of the 19 species exceeding 70 cm of maximum length were not sampled (i.e. 73%). Although nearly 45% of the largest species (i.e. >70 cm) for euryhaline and amphidromous families were not sampled, the taxonomic coverage was more evenly distributed across size classes, excepting the 80- to 90-cm size class with a single species that was not sampled (Fig. 3C). The influence of species size on taxonomic coverage, however, was not significant for either euryhaline and amphidromous families ( $\chi^2 = 6.22$ ;  $P = 0.79$ ) or primary freshwater families ( $\chi^2 = 11.38$ ;  $P = 0.33$ ), despite that more heterogeneity was observed for the later.

Amongst the 74 new records for both Java and Bali, a total of 20 introduced species belonging to 16 genera and eight families were newly recorded during the present campaign (Table 4). All of them displayed a single BIN with very low maximum intraspecific distances, excepting for *Clarias gariepinus* with two BINs diverging by 2.03% and *Poecilia reticulata* with two BINs diverging by 4.77%. The families Cichlidae and Poeciliidae were the



**Fig. 2** Distribution of genetic distances below and above species boundaries. (A) Distribution of maximum intraspecific distances (K2P). (B) Distribution of nearest neighbour distances (K2P). (C) Relationship between maximum intraspecific and nearest neighbour distances. Points above the diagonal line indicate species with a barcode gap.

most speciose with eight and five species, respectively. Amongst the 20 introduced species, four species originate from South America, nine from North and Central America, three from Africa and four from Asia.

## Discussion

The present study provides a DNA barcode reference library consisting of 159 species, including 139 native and 20 exotic species. Although maximum intraspecific distances exceeded nearest neighbour distances in a species pair (i.e. *Eleotris fusca* and *E. melanosoma*), all the species examined consisted of a diagnostic set of DNA barcodes arranged into monophyletic units. This study revealed that nearest neighbour distances were 28-fold higher on average than intraspecific maximum distances, suggesting that the present library is useful for the identification of Javanese and Balinese freshwater fishes. The average distance among congeneric, however, was 18-fold higher than the average intraspecific distance (0.57% vs. 10.21%), a range lower than that observed for maximum intraspecific and nearest neighbour distances. This difference was not unexpected considering the geographic focus of the present study. Nearest neighbour distances were estimated based on a geographically focused sampling. Considering that the Java and Bali islands include 227 species of the 900 species in Sundaland, the closest phylogenetic relatives have been probably sampled in only a few cases, thereby inflating estimates of nearest neighbour distances. This bias is not unexpected considering that the Pleistocene eustatic fluctuations repeatedly connected Sundaland major islands and offered temporary opportunities of dispersal during times of low sea levels and subsequent fragmentation during high sea level periods (Voris 2000; Woodruff 2010). As a consequence, patterns of allopatric speciation and trans-islands sister-species relationships have been frequently described (Dodson *et al.* 1995; Nguyen *et al.* 2008; Pouyaud *et al.* 2009; De Bruyn *et al.* 2013).

Although the barcoding gap was large in many species, expanding the spatial coverage of the present library will yield an increasing amount of closely related species, calling for its reassessment based on a more comprehensive spatial and taxonomic sampling. Along the same line, expanding the spatial coverage may be expected to reassess upward maximum intraspecific distances as previously demonstrated (Bergsten *et al.* 2012), particularly in fragmented landscapes (Geiger *et al.* 2014). Considering that very large maximum intraspecific distances were observed at the island level, even in numerous species consisting of a single BIN, the bias towards underestimated intraspecific distances may be high when incorporating a more comprehensive coverage of the distribution range for species straddling across the islands of Sundaland. The extent of this bias, however, is tightly linked to the dynamic that generated these high levels of intraspecific genetic diversity. If intraspecific divergence results from admixtures after secondary contacts of ancient populations previously isolated in allopatry, the bias may be limited. If intraspecific divergence results from the isolation of populations throughout fragmented landscapes, this bias may be expected to be high. Considering the intricate history of merging and isolation of populations, as a consequence of the Pleistocene eustatic fluctuations (Voris 2000; Woodruff 2010), both dynamics are likely to be involved and warrant further assessment of the barcoding gap at larger spatial scales in Sundaland. This hypothesis happens to be likely considering that the present study revealed 18 species with several BINs displaying a large extent of maximum divergence (1.2–12.81%).

The present study provided surprising results if considering the taxonomic coverage. Amongst the 227 species previously reported from Java and Bali, only 36% and 60% were sampled, respectively. This observation is dramatically contrasted by the 74 new records, among which 54 likely correspond to native species. Considering that most native species in Java and Bali correspond

**Table 2** Summary statistics of the 18 species with more than a single BIN including the number of BIN, maximum intraspecific distance and BIN accession numbers

No.	Family	Species	No. of BIN	Maximum distance (K2P)	Sympatric	BIN
1.	Gobiidae	<i>Acentrogobius caninus</i>	2	1.58		BOLD:ACE4188 BOLD:ACT1041
2.	Cyprinidae	<i>Barbodes binotatus</i>	3	5.13		BOLD:ACP5712 BOLD:ACP6025 BOLD:ACP6290
3.	Eleotridae	<i>Belobranchus belobranchus</i>	2	3.82		BOLD:ACQ5484 BOLD:ACQ5485
4.	Channidae	<i>Channa gachua</i>	5	5.63		BOLD:ACQ0290 BOLD:ACQ0291 BOLD:ACQ0292 BOLD:ACQ6939 BOLD:ACQ6940
5.	Clariidae	<i>Clarias gariepinus</i>	2	2.03		BOLD:AAB2256 BOLD:ACF4787
6.	Hemiramphidae	<i>Dermogenys pusilla</i>	3	2.28		BOLD:ACH7708 BOLD:ACH7709 BOLD:ACT1438
7.	Eleotridae	<i>Eleotris fusca</i>	2	12.81		BOLD:AAF0108 BOLD:ACQ5280
8.	Eleotridae	<i>Eleotris melanosoma</i>	2	6.2		BOLD:AAF0109 BOLD:AAK9481
9.	Gerreidae	<i>Gerres filamentosus</i>	2	10.83		BOLD:AAC0381 BOLD:AAY1477
10.	Eleotridae	<i>Giuris margaritacea</i>	2	12.56		BOLD:AAV6427 BOLD:ACP9929
11.	Sisoridae	<i>Glyptothorax platypogon</i>	5	5.99		BOLD:AAY1028 BOLD:ACP5850 BOLD:ACP5898 BOLD:ACP6223 BOLD:ACP6224
12.	Cobitidae	<i>Lepidocephalichthys hasselti</i>	2	3.25		BOLD:ACT2693 BOLD:ACT6514
13.	Mastacembelidae	<i>Macrognathus maculatus</i>	4	3.87		BOLD:ACT1648 BOLD:ACT1649 BOLD:ACT1650 BOLD:ACT1890
14.	Synbranchidae	<i>Monopterus albus</i>	2	7.83		BOLD:AAF8880 BOLD:ACT5080
15.	Adrianichthyidae	<i>Oryzias javanicus</i>	2	2.71		BOLD:ACT2454 BOLD:ACT6896
16.	Gobiidae	<i>Periophthalmus argentilineatus</i>	2	5.82		BOLD:AAY1920 BOLD:ACQ9240
17.	Cyprinidae	<i>Rasbora lateristriata</i>	2	1.79		BOLD:ACQ7159 BOLD:ACQ7160
18.	Osphronemidae	<i>Trichopodus trichopterus</i>	2	1.2		BOLD:AAE8555 BOLD:AAW0021

to old descriptions (Table S1, Supporting information) and that no comprehensive inventories of the Java and Bali ichthyofauna have been conducted during the last decade, new records were not unexpected, particularly for exotic species.

The international trade in ornamental fishes has grown rapidly over the last decades, as well as the

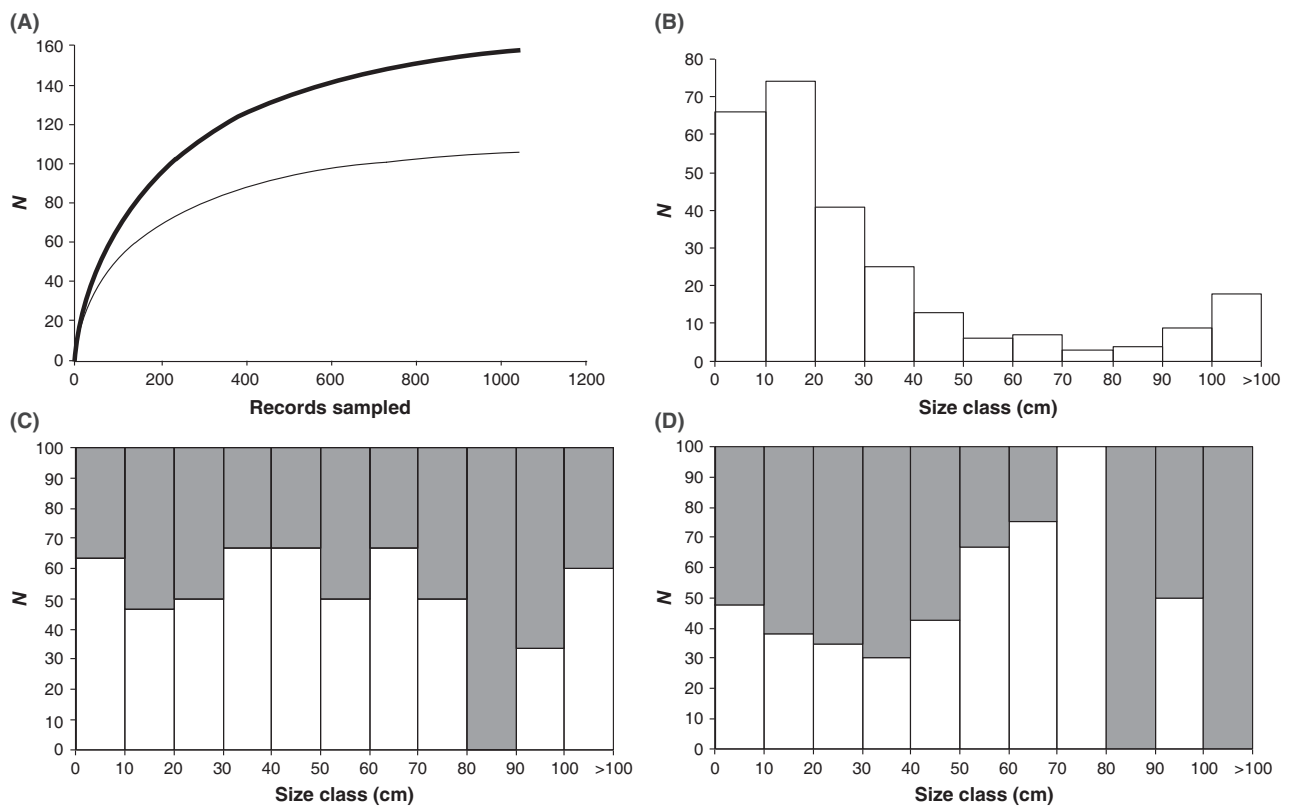
aquaculture trade, resulting in a global homogenization of the world ichthyofauna (Leprieur *et al.* 2007; Blanchet *et al.* 2010). The exotic species observed during the present campaign are among the most common of the ornamental (e.g. *Xiphophorus* spp., *Poecilia* spp., *Paraneotroplus* spp., *Archocentrus nigrofasciatus*, *Puntigrus tetrazona*) or aquaculture trade (e.g. *Oreochromis* spp., *Pangasianodon*

**Table 3** Summary statistics per families of the taxonomic coverage yielded by this study including the number of species derived from online catalogues, DNA barcoding coverage and new records

Family	N species from checklist	N species in Java			N species in Bali			Total	
		Checklist	Covered (%)	New records	Checklist	Covered (%)	New records	N species	Percentage of overage
Adrianichthyidae	2	2	1 (50)	0	1	1 (100)	0	2	50
Akysidae	3	3	0 (0)	0	0	N/A	0	3	0
Ambassidae	3	3	0 (0)	3	0	N/A	2	7	57
Anabantidae	0	0	N/A	1	0	N/A	0	1	100
Anguillidae	3	3	2 (67)	0	1	0 (0)	1	3	67
Antennariidae	1	1	0 (0)	0	0	N/A	0	1	0
Aplocheilidae	1	1	1 (100)	0	1	1 (100)	0	1	100
Apogonidae	2	2	0 (0)	0	1	0 (0)	0	2	0
Ariidae	3	3	1 (33)	0	0	N/A	0	3	33
Bagridae	9	9	4 (44)	0	0	N/A	0	9	44
Balitoridae	4	4	0 (0)	0	0	N/A	0	4	0
Belonidae	0	0	N/A	2	0	N/A	0	2	100
Blenniidae	1	1	0 (0)	0	0	N/A	0	1	0
Carangidae	0	0	N/A	1	0	N/A	0	1	100
Chanidae	0	0	N/A	1	0	N/A	0	1	100
Channidae	4	4	2 (50)	0	3	2 (67)	0	4	50
Cichlidae	0	0	N/A	8	0	N/A	2	8	100
Clariidae	3	3	1 (33)	1	0	N/A	0	4	50
Cobitidae	8	8	2 (25)	1	0	N/A	0	9	44
Cyprinidae	44	43	14 (33)	6	3	3 (100)	0	49	41
Dasyatidae	1	1	0 (0)	0	0	N/A	0	1	0
Eleotridae	11	10	6 (60)	4	4	4 (100)	5	15	73
Engraulidae	2	2	0 (0)	2	1	0 (0)	0	4	50
Gerreidae	1	1	1 (100)	1	1	1 (100)	0	2	100
Gobiidae	38	30	17 (57)	18	19	10 (53)	14	54	70
Haemulidae	1	1	0 (0)	1	0	N/A	0	2	50
Helostomatidae	0	0	N/A	1	0	N/A	0	1	100
Kuhliidae	2	2	1 (50)	0	0	N/A	0	2	50
Latidae	0	0	N/A	1	0	N/A	0	1	100
Leiognathidae	1	1	0 (0)	1	0	N/A	0	2	50
Loricariidae	0	0	N/A	1	0	N/A	0	1	100
Lutjanidae	1	1	1 (100)	0	0	N/A	0	1	100
Mastacembelidae	3	3	2 (67)	1	0	N/A	0	4	75
Megalopidae	1	1	0 (0)	0	0	N/A	0	1	0
Monodactylidae	1	1	0 (0)	0	0	N/A	0	1	0
Mugilidae	6	6	0 (0)	0	0	N/A	0	6	0
Muraenidae	2	2	0 (0)	0	0	N/A	0	2	0
Nandidae	1	1	0 (0)	0	0	N/A	0	1	0
Nemacheilidae	2	2	2 (100)	0	0	N/A	0	2	100
Notopteridae	3	3	1 (33)	0	0	N/A	0	3	33
Ophichthidae	1	1	0 (0)	2	0	N/A	2	4	75
Osphronemidae	4	4	3 (75)	1	0	N/A	1	5	80
Pangasiidae	2	2	0 (0)	2	0	N/A	0	4	50
Platycephalidae	0	0	N/A	1	0	N/A	0	1	100
Plotosidae	1	1	1 (100)	1	0	N/A	0	2	100
Poeciliidae	0	0	N/A	4	0	N/A	4	5	100
Pristidae	2	2	0 (0)	0	0	N/A	0	2	0
Pristigasteridae	2	2	0 (0)	0	0	N/A	0	2	0
Pristolepididae	1	1	0 (0)	0	0	N/A	0	1	0
Rhacichthyidae	1	1	1 (100)	0	1	1 (100)	0	1	100
Scatophagidae	1	1	1 (100)	0	0	N/A	0	1	100
Schilbeidae	1	1	1 (100)	0	0	N/A	0	1	100
Sciaenidae	4	4	2 (50)	0	0	N/A	0	4	50

Table 3 (Continued)

Family	N species from checklist	N species in Java			N species in Bali			Total	
		Checklist	Covered (%)	New records	Checklist	Covered (%)	New records	N species	Percentage of overage
Serrasalminidae	0	0	N/A	1	0	N/A	0	1	100
Siganidae	0	0	N/A	2	0	N/A	0	2	100
Sillaginidae	2	2	0 (0)	2	1	0 (0)	0	4	50
Siluridae	11	11	0 (0)	0	0	N/A	0	11	0
Sisoridae	2	2	1 (50)	0	0	N/A	0	2	50
Sparidae	1	1	0 (0)	0	0	N/A	0	1	0
Synbranchidae	1	1	1 (100)	0	0	N/A	1	1	100
Syngnathidae	8	7	3 (43)	2	1	1 (100)	2	9	56
Terapontidae	3	3	1 (33)	0	0	N/A	0	3	33
Tetraodontidae	4	4	0 (0)	1	0	N/A	0	5	20
Tetrarogidae	3	3	1 (33)	0	2	0 (0)	0	3	33
Toxotidae	1	1	1 (100)	0	0	N/A	0	1	100
Zenarchopteridae	3	3	1 (33)	1	0	N/A	0	4	50
Total	227	216	77 (36)	75	40	24 (60)	34	301	53



**Fig. 3** Accumulation curves and size class distributions of the 159 species analysed in this study. (A) Accumulation curves recovered from 100 iterations for species (bold curve) and genera (regular curve). (B) Distribution of size class (10 cm) of 266 species with documented maximum sizes among the 301 species of Java and Bali. (C) Percentages of species sampled (white) and not sampled (grey) across size class of 10 cm among the 164 species with documented maximum length for euryhalin or amphidromous families. (D) Percentages of species sampled (white) and not sampled (grey) across size class of 10 cm among the 135 species with documented maximum length for primary freshwater families.



**Table 4** Summary statistics of the 20 exotic species including the number of BIN, maximum intraspecific distance, geographic origin and BIN accession numbers

No.	Family	Species	No. of BIN	Maximum distance (K2P)	Origin	BIN
1.	Cichlidae	<i>Amphilophus citrinellus</i>	1	–	Central America (Nicaragua, Costa Rica)	BOLD:AAA7015
2.	Cichlidae	<i>Andinoacara rivulatus</i>	1	0	South America (Ecuador, Peru)	BOLD:AAJ5733
3.	Cichlidae	<i>Archocentrus nigrofasciatus</i>	1	0	Central America (Panama)	BOLD:AAD6407
4.	Cichlidae	<i>Oreochromis mossambicus</i>	1	0	Africa (Mozambique, Malawi)	BOLD:AAA8511
5.	Cichlidae	<i>Oreochromis niloticus</i>	1	–	Africa (widely distributed)	BOLD:ACR5811
6.	Cichlidae	<i>Parachromis managuensis</i>	1	–	Central America (Honduras, Costa Rica)	BOLD:AAB8227
7.	Cichlidae	<i>Paraneetroplus fenestratus</i>	1	–	North America (Mexico)	BOLD:AAD2571
8.	Cichlidae	<i>Paraneetroplus maculicauda</i>	1	–	Central America (Guatemala, Panama)	BOLD:AAB9907
9.	Clariidae	<i>Clarias gariepinus</i>	2	2.03	Africa (widely distributed)	BOLD:AAB2256 BOLD:ACF4787
10.	Cyprinidae	<i>Ctenopharyngodon idella</i>	1	0.16	Asia (China, Russia)	BOLD:ACL1923
11.	Cyprinidae	<i>Puntigrus tetrazona</i>	1	0	Asia (Indonesia: Sumatra, Borneo)	BOLD:AAD9761
12.	Loricariidae	<i>Liposarcus pardalis</i>	1	0	South America (Brazil, Peru, Bolivia)	BOLD:ACK1995
13.	Osphronemidae	<i>Trichopodus pectoralis</i>	1	0	Asia (Vietnam, Thailand, Cambodia)	BOLD:AAE8555
14.	Pangasiidae	<i>Pangasianodon hypophthalmus</i>	1	0	Asia (Vietnam, Thailand, Cambodia)	BOLD:AAE3237
15.	Poeciliidae	<i>Gambusia affinis</i>	1	0.72	North and Central America (USA, Mexico)	BOLD:AAC2756
16.	Poeciliidae	<i>Poecilia latipinna</i>	1	0	North America (USA, Mexico)	BOLD:ACE4147
17.	Poeciliidae	<i>Poecilia reticulata</i>	2	4.77	South America (widely distributed)	BOLD:ACC0443 BOLD:ACE3484
18.	Poeciliidae	<i>Xiphophorus hellerii</i>	1	0.16	North and Central America (Mexico to Honduras)	BOLD:AAB8020
19.	Poeciliidae	<i>Xiphophorus maculatus</i>	1	0	North and Central America (USA, Mexico)	BOLD:AAB7239
20.	Serrasalmidae	<i>Colossoma macropomum</i>	1	0.16	South America (widely distributed)	BOLD:AAD6423

*hypophthalmus*, *Clarias gariepinus*). Considering that Indonesia rank as one of the major contributors of the international ornamental fish trade (Ling & Lim 2005), this result highlights that avoiding escaping from fish farms or releasing in nature by fish hobbyist is challenging, particularly for species with high adaptive abilities in terms of breeding such as *Oreochromis* spp., *Xiphophorus* spp. or *Poecilia* spp. This situation is further amplified by the lack of knowledge by local populations of the geographic origin of most ornamental fish species traded in Indonesia and the impact of artificial introduction of exotic species. During the present survey, restocking with exotic species for recreational purposes was frequently observed in both Java (e.g. Lake Rawa Pening) and Bali (e.g. Lake Batur).

The observation of 54 new occurrences of species with no obvious economic interest for the ornamental trade is more challenging in terms of basic knowledge about the taxonomic composition of the Java and Bali ichthyofauna. This is particularly evident considering that this high number of new occurrences is drastically contrasted by the poor taxonomic coverage of the present campaign for the previously reported species according to online catalogues. Identifications were performed independently by several of the co-authors based on available monographs and later cross-validated, including based on DNA barcode data available in BOLD. Misidentifications are unlikely to account alone for this apparent discrepancy, an assertion further supported by the asymptotic trend of the accumulation curve for both

species and genera. Among the 16 species endemic of Java, for instance, only three were sampled, including *Nemacheilus chrysolaimos*, *Rasbora aprotaenia* and *Sicyopterus parvei*. The status of these endemic species should be revised, however, as *S. parvei* have been observed also in Bali. By contrast, the two endemic species of Bali were sampled (i.e. *Lentipes whittenorum* and *Rasbora baliensis*), as well as the three species endemic of Java and Bali (i.e. *Lentipes ikeae*, *Sycopus rubicundus* and *Stiphodon aureofuscus*). The validity of several endemic species of Java has been previously discussed as a substantial proportion of them correspond to old descriptions based on a single specimen such as *Barbodes platysoma*, *Mystus abbreviatus*, *Ompok javanesis*, *Puntius aphyia* and *Puntius bramoides* (Kottelat 2013; Eschmeyer & Fricke 2014). The contrasted coverage of the campaigns conducted in Java and Bali further questions their validity as sampling has been conducted nearby the type locality for several of them. Nomenclatural issues, however, are not sufficient *per se* to account for the large gap in the present taxonomic coverage, and extirpations are also likely considering the exponential increase of human populations, and associated anthropogenic perturbations, in Java. Despite that the influence of species maximum size on taxonomic coverage was not significant, for either euryhalin and amphidromous or primary freshwater families, many large-sized and emblematic primary freshwater species are missing from Java, as for instance *Pangasius djambal* (90 cm of SL), *Chitala chitala* (122 cm of SL), *Chitala lopis* (150 cm of SL), *Bagarius bagarius* (200 cm of TL), *Wallago attu* (240 cm of TL), *Tor tambra* (100 cm of TL) and *Tor soro* (100 cm of TL). Considering that several inland fisheries have been visited several times during the course of the study, including the Mojokerto fish market aggregating fisheries of the largest Javanese river (i.e. Brantas River), extirpations as a consequence of overexploitation cannot be discarded. Worth mentioning, the presence of most of these large species in Java is derived from the early ichthyological exploration of Javanese inland waters several decades or centuries ago (Roberts 1993).

## Conclusions

The present study highlights the difficulty to develop accurate DNA barcode reference libraries in Sundaland. Despite two centuries of ichthyological exploration of the inland waters of Java, the build-up of the taxonomic knowledge has been scattered in the literature, and as a consequence, the taxonomy of Javanese fishes has been accessible to only but a few specialists (Kottelat 2013). Considering the large discrepancies between the checklist established here and those available on online catalogues, the inventory of Javanese fishes is still far from

comprehensive and is currently plagued by uncertainties in the validity and occurrence of many species. The development of this DNA barcode library will allow more researchers to explore the ichthyodiversity in this part of Sundaland, what will without doubt help refine the checklist of Javanese fishes and probably shed a new light on the validity of the endemic species of Java. The Sundaland ichthyofauna is one of the world's most endangered and establishing accurate checklists for its major islands is an absolute priority. This library is a primer to that end.

## Acknowledgements

The authors wish to thank Dr. Siti Nuramaliati Prijono, Dr. Witjaksono, Mohammad Irham M.Sc., Dr. Marlina Adriyani, Dr. Rosichon Ubaidillah, Dr. Sri Sulandari, at Research Centre for Biology (RCB-LIPI), Dr. Jean-Paul Toutain, Dr. Jean-François Agnès and Dr. Domenico Caruso from the 'Institut de Recherche pour le Développement', Dr. Joel LeBail at the French embassy in Jakarta, Dr. Bambang Suryobroto and Dr. Achmad Farajallah at the Bogor Agronomy University for their support as well as Dr. Lukas Rüber and Dr. Fabian Herder for constructive discussions. We are thankful to Dr. Daisy Wowor and Dr. Daniel Lumbantobing at RCB-LIPI, Sumanta and Bambang Dwisusilo at IRD Jakarta and Dr. Helen Larson from the Museum and Art Gallery of the Northern Territory for their help during either the field sampling or the identifications of the specimens collected in Java and Bali. Part of the present study was funded by the Institut de Recherche pour le Développement (UMR226 ISE-M and IRD headquarter through incentive funds), the MNHN (UMR BOREA), the RCB-LIPI, the French Ichthyological Society (SFI), the Foundation de France and the French embassy in Jakarta. The Indonesian Ministry of Research approved this study, and field sampling was conducted according to the research permits 097/SIP/FRP/SM/IV/2014 for Philippe Keith, 60/EXT/SIP/FRP/SM/XI/2014 for Frédéric Busson and 41/EXT/SIP/FRP/SM/VIII/2014 for Nicolas Hubert. Sequence analysis was aided by funding from the Government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life project. This publication has ISEM number 2015-204.

## References

- Bergsten J, Bilton DT, Fujisawa T *et al.* (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, **61**, 851–869.
- Blagoev GA, deWaard JR, Ratnasingham S *et al.* (2016) Untangling taxonomy: a DNA barcode reference library for Canadian spiders. *Molecular Ecology Resources*, **16**, 325–341.
- 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.
- Clements R, Sodhi NS, Schilthuisen M, Ng PKL (2006) Limestone karsts of Southeast Asia: imperiled arks of biodiversity. *BioScience*, **56**, 733–742.
- Clements R, Ng PKL, Lu XX *et al.* (2008) Using biogeographical patterns of endemic land snails to improve conservation planning for limestone karsts. *Biological Conservation*, **141**, 2751–2764.

- De Bruyn M, Rüber L, Nylander S *et al.* (2013) Paleo-drainage basin connectivity predicts evolutionary relationships across three Southeast Asian biodiversity hotspots. *Systematic Biology*, **62**, 398–410.
- Dodson JJ, Colombani F, Ng PKL (1995) Phylogeographic structure in mitochondrial DNA of a South-east Asian freshwater fish, *Hemibagrus nemurus* (Sluroidei; Bagridae) and Pleistocene sea-level changes on the Sunda shelf. *Molecular Ecology*, **4**, 331–346.
- Eschmeyer WN, Fricke R (2014) *Catalog of fishes electronic version (updated 6 October 2014)*. San Francisco.
- Fraser S (2006) Threats to biodiversity. In: *The Ecology of Indonesia* (eds Marshall AJ, Beehler BM), pp. 1199–1229. Periplus Editions, Singapore.
- Froese R, Pauly D (2011) *Fishbase*. <http://www.fishbase.org>
- Geiger MF, Herder F, Monaghan MT *et al.* (2014) Spatial heterogeneity in the Mediterranean Biodiversity Hotspot affects barcoding accuracy of its freshwater fishes. *Molecular Ecology Resources*, **14**, 1210–1221.
- Hoffman M, Hilton-Taylor C, Angulo A *et al.* (2010) The impact of conservation on the status of the world's vertebrates. *Science*, **330**, 1503–1509.
- Hubert N, Hanner R (2015) DNA barcoding, species delineation and taxonomy: a historical perspective. *DNA Barcodes*, **3**, 44–58.
- Hubert N, Meyer C, Bruggemann JH *et al.* (2012) Cryptic diversity in Indo-Pacific coral reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. *PLoS ONE*, **7**, e28987.
- Hubert N, Kadarusman, Wibowo A *et al.* (2015) DNA barcoding Indonesian freshwater fishes: challenges and prospects. *DNA Barcodes*, **3**, 144–169.
- Ivanova NV, Zemlak TS, Hanner RH, Hébert PDN (2007) Universal primers cocktail for fish DNA barcoding. *Molecular Ecology Notes*, **7**, 544–548.
- Jaafar TNAM, Taylor MI, Mohd Nor SA, De Bruyn M, Carvalho GR (2012) DNA barcoding reveals cryptic diversity within commercially exploited Indo-Malay carangidae (Teleostei: Perciformes). *PLoS ONE*, **7**, e49623.
- Kadarusman, Hubert N, Hadiaty RK *et al.* (2012) Cryptic diversity in Indo-Australian rainbowfishes revealed by DNA Barcoding: implications for conservation in a biodiversity hotspot candidate. *PLoS ONE*, **7**, e40627.
- Keith P, Marquet G, Lord C, Kalfatak D, Vigneux E (2010) *Poissons et Crustacés D'eau Douce du Vanuatu*. Société Française d'Ichtyologie, Paris, France.
- Keith P, Marquet G, Gerbeaux P, Vigneux E, Lord C (2013) *Poissons et Crustacés D'eau Douce de Polynésie*. Société Française d'Ichtyologie, Paris, France.
- Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **15**, 111–120.
- Kottelat M (2013) The fishes of the inland waters of Southeast Asia: a catalog and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. *The Raffles Bulletin of Zoology*, **27**(Supplement), 1–663.
- Kottelat M, Whitten AJ, Kartikasari SR, Wirjoatmodjo S (1993) *Freshwater Fishes of Western Indonesia and Sulawesi*. Periplus Editions, Singapore.
- Lamoureux JF, Morrison JC, Ricketts TH *et al.* (2006) Global tests of biodiversity concordance and the importance of endemism. *Nature*, **440**, 212–214.
- Larson HK (2009) Review of the gobiid fish genera Egnathogobius and Pseudogobiopsis (Gobioidei: Gobiidae: Gobionellinae), with descriptions of three new species. *The Raffles Bulletin of Zoology*, **57**, 127–181.
- Larson HK (2010) A review of the gobiid fish genus Redigobius (Teleostei: Gobionellinae); with descriptions of two new species. *Ichthyological Exploration of Freshwaters*, **21**, 123–191.
- Leprieur F, Beauchard O, Blanchet S, Oberdorff T, Brosse S (2007) Fish invasions in the world's river systems: when natural processes are blurred by human activities. *Plos Biology*, **6**, 404–410.
- Ling KH, Lim LY (2005) The status of ornamental fish industry in Singapore. *Singapore Journal of Primary Industries*, **32**, 59–69.
- Lohman K, De Bruyn M, Page T *et al.* (2011) Biogeography of the Indo-Australian archipelago. *Annual Review of Ecology, Evolution and Systematics*, **42**, 205–226.
- Meyer C, Pauly G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, 2229–2238.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Nguyen TTT, Na-Nakorn U, Sukmanomon S, ZiMing C (2008) A study on phylogeny and biogeography of mahseer species (Pisces: Cyprinidae) using sequences of three mitochondrial DNA gene regions. *Molecular Phylogenetics and Evolution*, **48**, 1223–1231.
- Normile D (2010) Saving forests to save biodiversity. *Science*, **329**, 1278–1280.
- Pouyaud L, Sudarto, Paradis E (2009) The phylogenetic structure of habitat shift and morphological convergence in Asian Clarias (Teleostei, Siluriformes: Clariidae). *Journal of Zoological Systematics and Evolutionary Research*, **47**, 344–356.
- Rachmatika I (2003) *Fish fauna of the Gunung Halimun national park, West Java*. Biodiversity Conservation Project LIPJICA-PHKA, Jakarta, Indonesia.
- Ratnasingham S, Hebert PDN (2007) Bold: the barcode of life data system (<http://www.barcodinglife.org>). *Molecular Ecology Notes*, **7**, 355–364.
- Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS ONE*, **8**, e66213.
- Roberts TR (1993) The freshwater fishes of Java, as observed by Kuhl and van Hasselt in 1820–23. *Zoologische Verhandlungen*, **285**, 3–94.
- Schilthuizen M, Liew TS, Bin Elehan B, Lackman-Ancrenaz I (2005) Effects of karst forest degradation on pulmonate and prosobranch land snail communities in Sabah, Malaysian Borneo. *Conservation Biology*, **19**, 949–954.
- Smith AM, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society, Serie B*, **360**, 1825–1834.
- Sodhi NS, Koh LP, Clements R *et al.* (2010) Conserving southeast Asian forest biodiversity in human-modified landscapes. *Biological Conservation*, **143**, 2375–2384.
- Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153–1167.
- Woodruff DS (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugium-phase biodiversity. *Biodiversity and Conservation*, **19**, 919–941.

---

N.H., R.K.H., R.H., P.K. designed the study. F.B., H.D., A.H., S.S., N.H., R.K.H., P.K. performed the study. N.H., H.D. & A.H. analyzed the data. H.D., N.H. wrote the paper. N.H., P.K. & F.B. obtained sampling permit.

---

### Data accessibility

All collecting and sequence data are available on the Barcode of Life Datasystem (BOLD) in the projects 'Barcoding Indonesian Fishes – part II. Inland fishes of Java and Bali [BIFB]', 'Barcoding Indonesian Fishes – part III. Sicydiinae of Sundaland [BIFC]', 'Barcoding Indonesian Fishes – part VIb. Widespread primary freshwater fishes of Java and Bali [BIFGA]', 'Barcoding of Indonesian

Fishes – part VIIb *Rasbora* spp [BIFHB] in the container 'Barcoding Indonesian Fishes' of the 'Barcoding Fish (FishBOL)' campaign. The sequence alignment and neighbour-joining tree (as both PDF and Newick files) have all been uploaded to DRYAD (doi: 10.5061/dryad.tk5rj). The sequences are also available on GenBank (see Table S2, Supporting information for accession numbers).

### Supporting Information

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

**Figure S1** Midpoint rooted Neighbor-joining tree of the 1046 DNA barcodes collected from the 159 species analyzed in this study.

**Table S1** Checklist of the freshwater fishes of Java and Bali including the authors and date of the original description, maximum length, type of length measurement, status of the species namely native or introduced, source references for the distribution, status of the distribution including endemic, occurring in other countries or original distribution range for introduced species, occurrence in Java, potential new occurrence in Java, occurrence in Bali, potential new occurrence in Bali.

**Table S2** Collecting data and sequence information.

**Table S3** Barcoding gap in the species analyzed in the present study including mean intra-specific, maximum intra-specific and nearest neighbor distances for the 159 species analyzed in the present study.