Vol. 583: 179–193, 2017
https://doi.org/10.3354/meps12316

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published November 16

Geography and life history traits account for the accumulation of cryptic diversity among Indo-West Pacific coral reef fishes

Nicolas Hubert[1,*], Agnès Dettai[2], Patrice Pruvost[3], Corinne Cruaud[4], Michel Kulbicki[5], Robert F. Myers[6], Philippe Borsa[5]

[1]Institut de recherche pour le développement (IRD), UMR 226, Institut des sciences de l'évolution de Montpellier (ISEM), CNRS/IRD/UM/CIRAD, Montpellier, France

[2]Muséum national d'histoire naturelle (MNHN), Institut de systématique et evolution (ISYEB), MNHN/CNRS/EPHE/UPMC-UMR 7205, Sorbonne Universités, Paris, France

[3]Muséum national d'histoire naturelle (MNHN), UMR 207, Biologie des organismes et écosystèmes aquatiques (BOREA), MNHN/CNRS/UPMC/IRD, Paris, France

[4]Genoscope, CEA, DRF, IG, Evry, France

[5]Institut de recherche pour le développement (IRD), UMR 250 Ecologie marine tropicale des océans Pacifique et Indien/LabEx Corail, Nouméa, New Caledonia

[6]Seaclicks/Coral Graphics, Wellington, FL, USA

ABSTRACT: Indo-West Pacific coral reef fishes form speciose ecological communities. A biogeographically meaningful interpretation of diversity patterns in this region requires accurate inventories of species. Previous studies have suggested that biogeographic scenarios for Indo-West Pacific coral reef fishes are compromised by an unacknowledged yet substantial amount of cryptic diversity. DNA barcoding, the use of a mitochondrial gene as an internal species tag for species identification, has opened new perspectives on global biodiversity. The present study, based on the largest DNA barcode reference library produced to date for Indo-West Pacific coral reef fishes, sheds new light on the extent of cryptic diversity and its evolutionary origin. We analyzed 3174 DNA barcodes for 805 species of coral reef fishes sampled at 3 different locations across the Indo-West Pacific (including 538 new DNA barcodes for 270 species sampled from New Caledonia). Among the 183 species with Indo-West Pacific distribution and multiple specimens analyzed, 78 (42.6%) were represented by 2 or more monophyletic lineages alternatively sorted between the sampling sites in the Indian and Pacific oceans and another 73 (40%) showed evidence of phylogeographic structure. Spatial analyses pointed to a detectable impact of geographic isolation on the emergence of cryptic diversity. The significant correlation of several life history traits to the maximum intraspecific genetic distances suggests that genetic divergence among geographically isolated cryptic lineages accumulated though mutation and genetic drift.

KEY WORDS: Biogeography · Biodiversity · Evolution · Vicariance

INTRODUCTION

Because species are the fundamental components of ecosystems, delineating species is a mandatory step in the study of their ecological and evolutionary dynamics (Hubbell 2001, Ricklefs 2003, McPeek 2008, Brodersen & Seehausen 2014). Whatever the spatial scale under study, species are hypotheses (Hubert & Hanner 2015, Pante et al. 2015), and the accuracy of these hypotheses directly impacts our

*Corresponding author: nicolas.hubert@ird.fr

Publisher: Inter-Research · www.int-res.com

ability to infer eco-evolutionary processes from species distributions and abundances (Smith et al. 2007, 2008, Hubert et al. 2012, Tornabene et al. 2015). The increased use of molecular approaches for species delineation and identification, associated with the development of the international Barcode of Life project, has highlighted biases in our perception of species richness (Hebert et al. 2004, Smith et al. 2007, 2008, Hubert et al. 2012). DNA barcoding, the use of a single mitochondrial gene as a universal internal species tag for species identification, has opened up new perspectives in the inventory of Earth's biotas by streamlining the taxonomic workflow (Monaghan et al. 2009, Butcher et al. 2012, Tänzler et al. 2012, Riedel et al. 2013). The benefit of using standardized molecular approaches such as DNA barcoding for large-scale species inventories is particularly evident for speciose tropical biotas, where unexpectedly high levels of cryptic diversity have been abundantly documented (Hebert et al. 2004, Janzen et al. 2005, Smith et al. 2007, 2008, Hubert et al. 2012, Mat Jaafar et al. 2012, Kadarusman et al. 2012, Durand et al. 2017) and where large numbers of putative species have been concomitantly described through a fast standardized approach involving DNA barcoding (Butcher et al. 2012, Riedel et al. 2013).

Coral reefs represent an example of highly diverse and complex ecosystems where the use of molecular species identifications has opened new perspectives in evolutionary and ecological dynamics (Meyer & Paulay 2005, Barber & Boyce 2006, Bucklin et al. 2011, Plaisance et al. 2011, Hubert et al. 2012, 2015, Mat Jaafar et al. 2012, Ko et al. 2013, Leray et al. 2013, Tornabene et al. 2015). Teleost fishes constitute the most diverse group of vertebrates in coral reefs, and recent studies have uncovered high levels of cryptic diversity at various spatial scales in Indo-West Pacific reef fishes, ranging from regional (i.e. ocean basin) to local (i.e. island) scales (Hubert et al. 2012, Winterbottom et al. 2014, Tornabene et al. 2015, Victor 2015). Unacknowledged cryptic diversity in Indo-West Pacific reef fish communities hampers our understanding of diversity patterns (Leprieur et al. 2012, Kulbicki et al. 2013, Mouillot et al. 2013, Pellissier et al. 2014) and their origin (McCafferty et al. 2002, Kuriiwa et al. 2007, Drew & Barber 2009, Leray et al. 2010, Winters et al. 2010, Gaither et al. 2011, Borsa et al. 2013, Sims et al. 2014, Durand & Borsa 2015, Randall & Victor 2015, Tornabene et al. 2015). A broad assessment of the factors promoting cryptic diversity in coral reef ecosystems is timely, because the exploration of diversity patterns in the Indo-West Pacific is an increasingly debated topic (Jokiel & Martinelli 1992, Briggs 1999, 2000, 2003, Halas & Winterbottom 2009, Hubert et al. 2012, Gaither & Rocha 2013, Tornabene et al. 2015). The conservation of coral reefs also requires up-to-date checklists of species whose accuracy depends on the approach involved in species delineation (Hubert et al. 2012, Winterbottom et al. 2014, Pante et al. 2015, Tornabene et al. 2015, Victor 2015).

Here we present a large-scale DNA barcoding study of coral reef fishes at 3 sites in the Indian and Pacific oceans to explore the influence of geographic isolation and life history traits (LHTs) on the emergence of cryptic diversity and to identify the factors that promote cryptic diversity in Indo-West Pacific coral reef fishes. Based on a DNA barcode reference library of 2276 sequences belonging to 668 species, Hubert et al. (2012) observed that cryptic diversity was characteristic of nearly half of the species with Indo-West Pacific geographic range. Cryptic lineages were genetically distant from their nearest neighbor by 1 to 12% nucleotide distance, which is substantially higher than the average values of maximum genetic distances reported for marine fish species (Ward et al. 2005, Victor et al. 2015). In the present study, we produced a new DNA barcode dataset from New Caledonia (NC) and added it to Hubert et al.'s (2012) revised dataset from the western Indian Ocean (WIO) and French Polynesia (FP). We analyzed 3174 DNA barcodes from 805 species of coral reef fishes to re-examine the extent of cryptic diversity within species with Indo-West Pacific or pan-Pacific distributions and to address the following questions: (1) Is geographic isolation a factor promoting the emergence of cryptic diversity? (2) Are LHTs related to generation time and dispersal ability related to the genetic distance among cryptic lineages?

MATERIALS AND METHODS

Building a DNA barcode reference library for Indo-West Pacific coral reef fishes

Large-scale DNA barcode surveys have been conducted on marine fishes in the Indian and Pacific oceans in the last decade (Hubert et al. 2010, 2011, 2012, 2015). The corresponding DNA barcode reference libraries are publicly available through the Barcode of Life Data System (BOLD; Ratnasingham & Hebert 2007). BOLD is an online repository dedicated to DNA barcode data. It features a workbench that enables the assembly and curation of DNA barcode reference libraries. In particular, BOLD allows users to tag and annotate specimen records and as-

sociated metadata, once these have been released publicly. This promotes open-access updates on specimen identification (e.g. changes in nomenclature, corrections of misidentifications), thus leading to continuous improvements of the taxonomic accuracy of DNA barcode records. Since the publication of the first DNA barcode reference libraries for Indo-West Pacific reef fishes (Hubert et al. 2010, 2011, 2012, 2015), several records have been tagged as misidentified or pending taxonomic updates in the BOLD projects IPCOM and SBF in the WIO, including Madagascar and Reunion (WIO), and MBFA, MBFC, FPFL and FPFLB in FP in the south central Pacific. Since the data produced by the foregoing projects have been made available in BOLD, 294 records from 71 species have been updated after re-examination of the voucher specimens or coloration patterns from their photographs. Among these, 196 updates (41 species) were nomenclatural, i.e. species resurrection, new species description, or assignment to a different genus; 10 updates (2 species) were new identifications; 81 updates (27 species) were corrections of misidentifications; and 7 updates (2 species) consisted of flagging low-quality sequences, hence excluding them from the BOLD search engine (Table S1 in the Supplement at www.int-res.com/articles/suppl/m583p179_supp.xls). As a result, 2636 records identified to species level (688 species, 284 genera) were analyzed.

In 2010, a new campaign named RESICOD targeted reef fishes in NC (Coral Sea, southwestern Pacific), with the aim to produce a DNA barcode reference library. During the RESICOD campaign, 609 specimens were referenced and ancillary information including sampling details, geographic coordinates and specimen photographs was deposited in BOLD. The specimens were collected from 5 to 7 December 2010 from a reef site in Baie Sainte-Marie (22° 19′ S, 166° 28′ E) in the southwestern lagoon of NC on the flat, crest, and external slope of the reef barrier near Dumbea Pass (22° 20′ S, 166° 14′ E), at depths ranging from ca. 1 to ca. 10 m. Specimens were caught using powdered derris root (containing about 7 to 8% rotenone) released within underwater quadrats each measuring about 20 × 20 m. All specimens were preserved in formalin, with a tissue subsample (i.e. fin clip or muscle biopsy) simultaneously preserved in ethanol for subsequent DNA sequencing. Specimens were deposited as vouchers at the ichthyological collections of the Muséum national d'histoire naturelle (MNHN), Paris, under registration nos. MNHN 2010-1074 to MNHN 2010-1822. Identifications were based on morphological criteria (color, meristic counts and proportional measurements) given in recent monographs including Food and Agriculture Organization of the United Nations (FAO) identification guides for fishery purposes (FAO, Rome; www.fao.org/) as well as regional identification guides (Randall et al. 1997, Myers 1999, Laboute & Grandperrin 2000, Randall 2005) and numerous individual papers in the recent scientific literature.

Sequencing DNA barcodes

Genomic DNA was extracted by an EPmotion 5075 extraction robot (Eppendorf AG) with the Macherey-Nagel 96 Tissue extraction kit following the manufacturer's instructions. A 650 bp segment from the 5 region of the cytochrome oxidase I gene was amplified using the primer pairs TelF1 (5′-TCG ACT AAT CAY AAA GAY ATY GGC AC-3′)/TelR1 (5′-ACT TCT GGG TGN CCA AAR AAT CAR AA-3′) (Dettai et al. 2011) or TelF1/FishR1 (5′-TAG ACT TCT GGG TGG CCA AAG AAT CA-3′) (Ward et al. 2005). PCR amplifications used 5% volume DMSO, 5 mg bovine serum albumin, 250 mM of each dNTP, 0.25 mM *Taq* DNA polymerase (Qiagen), 2 µl of the corresponding 10X buffer, 1.5 pM of each of the 2 primers and molecular biology grade water to a volume of 20 µl, with 3 µl of DNA extract. Amplification was achieved following heating at 94°C (5 min); 45 cycles of 94°C (30 s), 50°C (30 s) and 72°C (45 s); followed by a final extension at 72°C (3 min). The PCR products were purified and their nucleotide sequences determined using the method of Sanger at the Genoscope platform (Evry, France; www.genoscope.cns.fr/) using the same primers. Both strands were sequenced and checked against their chromatogram using Sequencher 4.8 (Gene Codes). All sequences were deposited in GenBank (www.ncbi.nlm.nih.gov/). Accession numbers for the barcodes, specimen and collection data, sequences, trace files and primer details are available in the RESIC project file under the general container 'Barcoding Fish (FishBOL)' in BOLD (www.barcodinglife.org). The collection data are available in Table S2 in the Supplement.

Spatial analyses of genetic divergence

Sequence divergence was estimated according to the Kimura 2-parameter (K2P) model of nucleotide substitution. The midpoint-rooted neighbor-joining (NJ) tree of K2P distances was built to provide a graphic representation of species divergence using the Sequence Analysis module of BOLD (Table S3

in the Supplement). Distributions of the average and maximum intraspecific K2P distances were first compared to the distribution of the K2P distances to the nearest neighbor. This step enabled us to detect a potential barcoding gap (i.e. the lack of overlap between the distribution of the maximum intraspecific divergences and the distribution of the distances to the nearest neighbor; Meyer & Paulay 2005). The distributions of the mean and maximum intraspecific K2P distances were estimated at each site (WIO, NC, FP) using class intervals of 1% width. The 6 distributions obtained were further compared, and the null hypothesis of independence between class intervals of genetic divergence and sites was tested using Fisher's exact test as implemented in the R Stats v. 3.1.2 package (R Core Team 2014).

Then, we examined the distribution of pairwise K2P distances at several spatial scales, ranging from sites (i.e. local) to several schemes of aggregated sites (i.e. regional), to assess the influence of geographic isolation and population fragmentation. Considering that isolated populations with limited or no gene flow will tend to accumulate genetic divergence through mutation and genetic drift, the distribution of intraspecific K2P distances is expected to be skewed toward higher values if aggregating sites are separated by larger geographic distances (Bergsten et al. 2012). To test this assumption, we aggregated sequences from 2 different sites (i.e. addressing the regional scale) and estimated the distribution of both mean and maximum intraspecific K2P distances for the species sampled at both sites. These distributions were compared to the distributions of the same distances estimated within a single site. Considering that intraspecific K2P distances were estimated based on limited sample sizes at the local scale (i.e. an average of 2 to 5 individuals per species per site), the distributions of the K2P distances within sites were established based on the maximum value from the 2 sites aggregated (e.g. maximum K2P distance of WIO and NC when comparing with WIO+NC). We compared the distributions of the mean and maximum intraspecific K2P distances at the regional scale (i.e. for WIO+NC and FP+NC) with the distributions of the highest mean and maximum intraspecific K2P distances at the local scale (i.e. WIO or NC, and FP or NC, respectively). The null hypothesis that the individual ranking is equivalent between the 2 distributions was tested with the non-parametric Wilcoxon signed rank test (*T*-test) for matched samples, as implemented in the R Stats v. 3.1.2 package (R Core Team 2014). Although the distributions of the maximum and mean intraspecific K2P distances were obtained using class intervals of 1% width of genetic distance, exact K2P distance values were used for the *t*-test.

LHTs and genetic divergence

The accumulation of genetic divergence among populations results from the balancing effects of mutation, natural selection, genetic drift and gene flow, the latter two being influenced by several LHTs such as generation time and dispersal (Kingman 1982, Hudson 1983, Kimura 1983, Tajima 1983). The balance between gene flow and genetic drift at neutral loci may be expected to vary according to the LHTs that are potentially associated with higher dispersal ability such as a long pelagic larval duration (PLD) and a large geographic range size (GRS). In a set of connected populations, population divergence may be positively correlated with geographic distance as expected under the isolation-by-distance model of population genetics. The slope of the relationship, however, depends on neighborhood size, itself linked to dispersal ability, and pace of genetic drift (i.e. generation time) (Rousset 1997). Alternatively, isolated populations are expected to accumulate divergence at a rate determined by the pace of genetic drift, itself determined by generation time and effective population size. We tested the null hypothesis that populations are geographically connected by examining the relationship between the LHTs related to dispersal and generation time and genetic distance among cryptic lineages. A negative correlation between dispersal and genetic divergence would support an isolation-by-distance model. By contrast, a negative correlation between generation time and genetic divergence would support a model of physical isolation without gene flow. Maximum age may be considered as a surrogate of generation time; however, this information is documented for only a handful of species (Froese & Pauly 2016). In this context, maximum standard length (SL) was used as a proxy of generation time by assuming that small species have generally shorter generation times than larger species on a broad taxonomic scale (Martin & Palumbi 1993, April et al. 2013).

Maximum intraspecific K2P distance, SL, PLD and GRS were compiled for the species distributed in more than a single location. SL, PLD and GRS were linearized after log transformation. The null hypothesis of no correlation between maximum intraspecific K2P distance and SL, PLD or GRS was tested using Pearson's R test as implemented in the R Stats v. 3.1.2

package (R Core Team 2014). GRS was estimated based on species occurrences across an Indo-Pacific 5° × 5° grid from Parravicini et al. (2014) and expressed in terms of the number of cells in which a species occurs. SL data were collected from FishBase (Froese & Pauly 2016), and PLD data were obtained from Luiz et al. (2013).

RESULTS

DNA barcoding gap, species identification and cryptic diversity

We failed to amplify 70 specimens, that is, 11.5% of the 609 specimens from the RESICOD campaign in NC. Thus, among the 609 reef fish specimens from NC, a full-length PCR product (652 bp) was obtained for 538 individuals of 270 species from 138 genera and 86 families. No insertion/deletion or stop codon was found, as expected in a functional protein-coding sequence. To this DNA barcode dataset from NC, we added the previously released sequences from the set of 2636 specimens identified to the species level from the projects IPCOM, SBF, MBFA, MBFC, FPFL and FPFLB (central Pacific). This yielded 3174 sequences from Indo-West Pacific coral reef fishes. These were assigned to 805 species under the current nomenclature (Table S2 in the Supplement). The K2P-NJ tree derived from the analysis of the complete sequence dataset is available as Appendix Fig. A1. Overall, sequences from 1196 individuals belonging to 378 species from the WIO, 1440 individuals belonging to 461 species from FP and 538 individuals belonging to 270 species from NC were included.

Table 1. Summary of Kimura 2-parameter (K2P) distances for increasing taxonomic levels. Data are from 3174 sequences from 805 Indo-West Pacific coral reef fish species (see Fig. A1 in the Appendix). Intraspecific K2P distances were based on the 604 species with N > 1. NC: New Caledonia; FP: French Polynesia: WIO: western Indian Ocean

	n	No. of taxa	No. of comparisons	K2P distance (%) Min.	Mean	Max.
Intraspecific						
WIO	1092	275	2041	0	0.44	20.02
NC	436	140	559	0	0.39	19.09
FP	1332	353	3853	0	0.35	23.36
WIO+NC	1566	374	3384	0	1.08	21.09
FP+NC	1828	450	5324	0	0.59	24.35
WIO+NC+FP	2977	609	10115	0	1.12	24.35
Among species within genus						
WIO	973	83	8062	0.31	14.97	27.13
NC	386	53	1647	5.32	15.8	29.84
FP	1021	85	10635	0.62	13.53	28.39
WIO+NC	1475	106	16403	0	14.97	29.84
FP+NC	1538	111	19192	0	14.2	29.84
WIO+NC+FP	2685	138	49903	0	14.4	31.57
Among genera within family						
WIO	1102	28	44633	5.79	20.23	30.2
NC	493	25	7480	2.04	20.47	30.52
FP	1317	34	38619	4.82	20.72	31.91
WIO+NC	1618	31	87367	1.88	20.29	31.05
FP+NC	1846	38	78931	2.04	20.65	31.91
WIO+NC+FP	3001	45	233662	1.87	20.43	33.1

A steady increase in genetic variation with increasing taxonomic level was observed (Table 1). Genetic divergence among congeneric species was on average 13 times higher than among conspecific individuals. Intraspecific divergence, however, showed considerable heterogeneity, ranging from 0 to 24.35% with a mean of 1.12%, a value 3 times higher than previously reported for fishes (Ward et al. 2005, 2008, April et al. 2011, Pereira et al. 2013). The distributions of mean intraspecific and maximum intraspecific distances largely overlapped the distribution of the distance to the nearest neighbor (Fig. 1), and no barcoding gap was observed. DNA barcodes did not reflect species boundaries in 6 species pairs, where sequences were mixed into a single cluster. These were *Abudefduf sexfasciatus* (N = 7 individuals) and *A. vaigiensis* (N = 2), *Chaetodon guttatissimus* (N = 6) and *C. pelewensis* (N = 5), *C. kleinii* (N = 6) and *C. trichrous* (N = 3), *C. interruptus* (N = 3) and *C. unimaculatus* (N = 3), *Cephalopholis nigripinnis* (N = 5) and *C. urodeta* (N = 3), and *Arothron meleagris* (N = 5) and *A. nigropunctatus* (N = 1). These cases represented only 1% of the 590 species sampled and sequenced for more than 1 individual. This estimate did not take into account the 206 species that were represented by only 1 individual; these may contain additional cases of mixed DNA barcode clusters. Cases of shared haplotypes between species were only observed at the regional level. Minimum distances of zero between congeneric species were only observed when pooling WIO+NC, NC+FP or WIO+NC+FP. By contrast, minimum distances between congeneric species ranged from 0.31 to 5.32 when WIO or NC or FP sites were considered individually (Table 1).

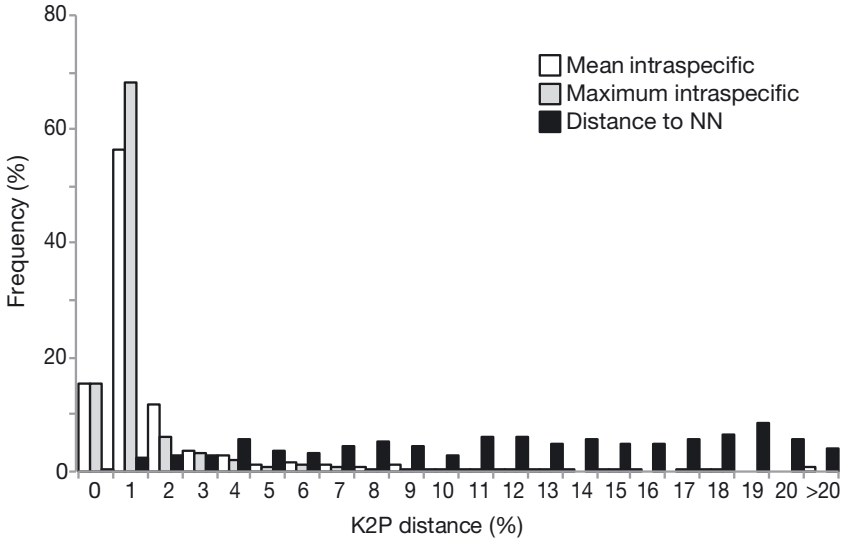

Fig. 1. Summary of the distribution of Kimura 2-parameter (K2P) distances to the nearest neighbor (NN), and mean and maximum intraspecific K2P distances

Influence of geography and LHT on population divergence

We inspected the distributions of intraspecific K2P distance (both mean and maximum) at various spatial scales (Table 1, Fig. 2). The distributions of the intraspecific K2P distances in the WIO, NC and FP exhibited similar trends (Fig. 2a). The independence of class interval of genetic divergence and sites was not rejected by Fisher's exact test for mean distances (WIO vs. NC: $p = 0.364$; WIO vs. FP: $p = 0.866$, FP vs. NC: $p = 0.669$) and maximum distances (WIO vs. NC: $p = 0.352$, WIO vs. FP: $p = 0.10$, FP vs. NC: $p = 0.356$). By contrast, the null hypothesis that the individual ranking of intraspecific K2P distances is equivalent between WIO or NC (Fig. 2a) and in the Indo-West Pacific (WIO+NC; Fig. 2b) was rejected (WIO or NC vs. WIO+NC; Wilcoxon signed rank test; $p < 0.001$ for both mean and maximum intraspecific distance). Comparisons within the Pacific Ocean provided more contrasted results. The null hypothesis that the individual ranking of mean intraspecific K2P distances is equivalent between FP or NC (Fig. 2a) and FP+NC (Fig. 2c) was not rejected (FP or NC vs. FP+NC: $p = 0.1633$). By contrast, the null hypothesis of equivalent distribution of the maximum intraspecific distance between FP or NC and FP+NC was rejected (FP or NC vs. NC+FP: $p < 0.0001$). The distributions of intraspecific K2P distances were skewed toward higher values when different sites were included (WIO+NC or NC+FP; Fig. 2b,c) and tended to overlap more with the distribution of the K2P distances to the nearest neighbor (Fig. 2d,e).

Among the 217 species with Indo-West Pacific distribution, 185 were represented by a sample of >1 individual at each of the 3 sampling sites. These species followed one of 3 patterns (Fig. 3, Table S3 in the Supplement). The first, which concerned 93 species, was a lack of spatial structure between populations from the WIO and populations from FP and NC (Fig. 4, Table S3), including 20 that had shared haplotypes (Fig. 4: pattern I.1) and 73 without shared haplotypes (Fig. 4: pattern I.2). The second pattern, found in 78 species, was that of substantial genetic difference (i.e. cryptic diversity) between populations from the WIO and populations from FP and NC (Table S4 in the Supplement). This included 64 species displaying reciprocal monophyly between oceans (Fig. 4: pattern II.1) and 14 that showed spatial paraphyly (Fig. 4: pattern II.2). The third pattern concerned 12 species harboring several deeply diverging lineages and which were either polyphyletic or paraphyletic from a taxonomic perspective (Fig. 4: pattern III).

We restricted the comparisons between maximum intraspecific K2P distances and LHTs to the species that showed clear patterns of allopatric divergence (Fig. 4: pattern II.1). Considering that taxonomic uncertainties and introgressive hybridization cannot be discarded for the polyphyletic and paraphyletic species belonging to pattern III, we excluded these species from the analysis of the relationship between K2P distances and LHT to avoid a bias toward overestimating maximum intraspecific K2P distances. The maximum intraspecific K2P distance of the widespread species that showed reciprocal monophyly between WIO and Pacific Ocean sites (Fig. 4: pattern II.1) was significantly correlated to SL ($r = -0.392$; $p = 0.001$) and GRS ($r = -0.272$; $p = 0.027$) but not significantly correlated to PLD ($r = -0.039$; $p = 0.772$).

DISCUSSION

DNA barcoding, species identification and cryptic diversity

Our study confirmed the potential of DNA barcodes for automated species identification in Indo-West Pacific coral reef fishes and illustrates the benefits of using a mitochondrial DNA barcode, such

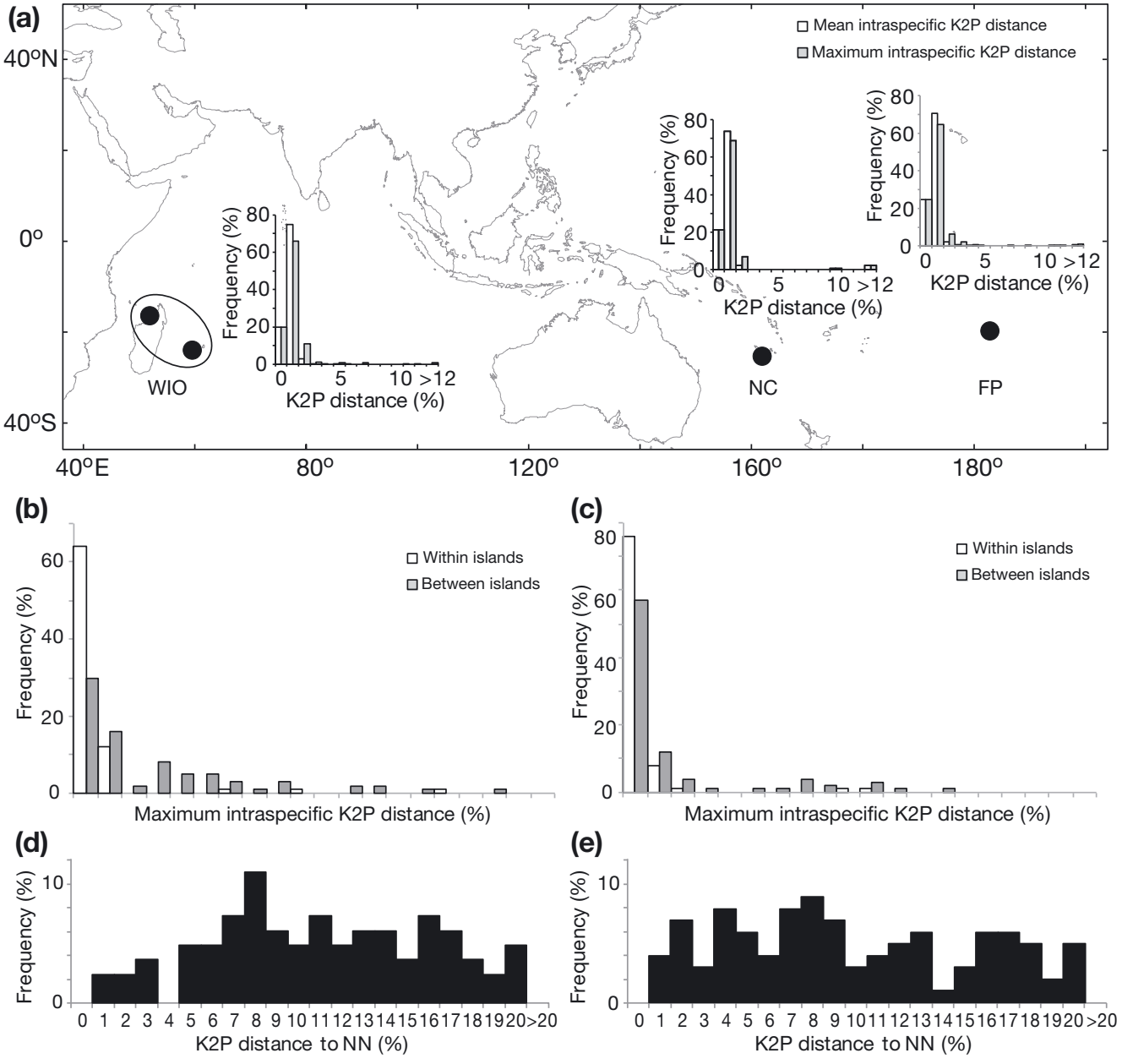

Fig. 2. Distributions of mean Kimura 2-parameter (K2P) distance, maximum K2P distance and K2P distance to the nearest neighbor (NN), within and among sites. (a) Sampling locations for the 3 DNA barcoding campaigns referred to in the present study, i.e. western Indian Ocean (WIO), New Caledonia (NC) and French Polynesia (FP), with respective distributions of intraspecific K2P distances (white indicates mean; grey indicates maximum). (b) Distributions of the maximum intraspecific K2P distances within WIO or NC (white) and for the pooled DNA barcode libraries from WIO and NC based on shared species only (grey). (c) Distributions of the maximum intraspecific K2P distances within FP or NC (white) and for the pooled DNA barcode libraries from NC and FP based on shared species only (grey). (d) Distribution of the K2P distances to NN in the pooled DNA barcode libraries of WIO and NC. (e) Distribution of the K2P distances to NN in the pooled DNA barcode libraries of NC and FP

as a high number of copies and higher rate of genetic drift and mutation compared to nuclear genes, and the ease of sequencing thanks to the availability of universal primers for fishes. Among the 590 species represented by more than a single individual, 580 (98%) were characterized by a monophyletic cluster of DNA barcodes (i.e. 1 or more lineages that are monophyletic and do not include any individuals of another species). DNA barcodes failed to reflect species boundaries for only 12 species (2% of 590). Nearly 99% of the successfully sequenced specimens in the complete dataset were reliably assigned to the

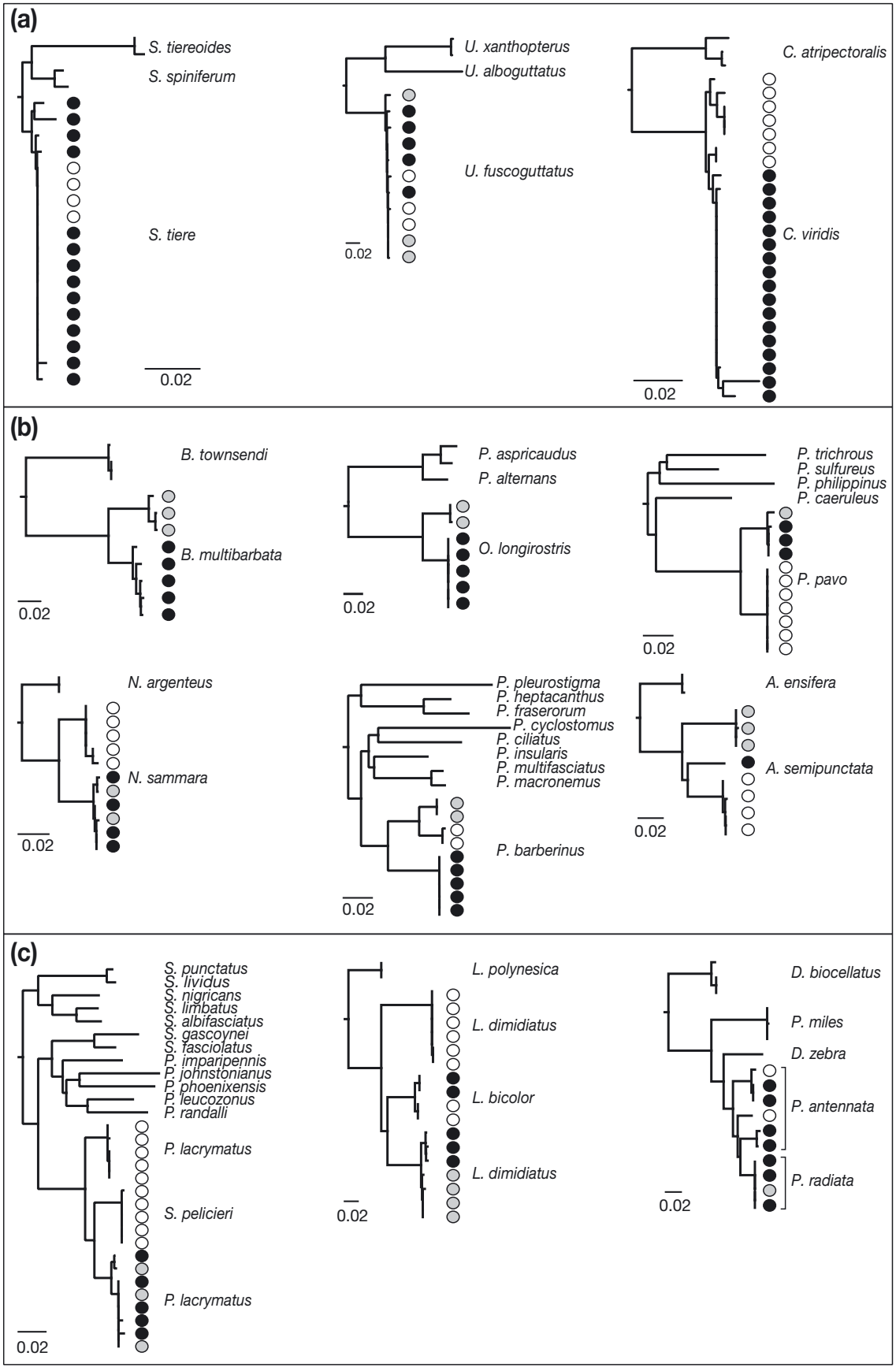
(a)
S. tiereoides
S. spiniferum
S. tiere
0.02
U. xanthopterus
U. alboguttatus
U. fuscoguttatus
0.02
C. atripectoralis
C. viridis
0.02
(b)
B. townsendi
B. multibarbata
0.02
P. aspricaudus
P. alternans
O. longirostris
0.02
P. trichrous
P. sulfureus
P. philippinus
P. caeruleus
P. pavo
0.02
N. argenteus
N. sammara
0.02
P. pleurostigma
P. heptacanthus
P. fraserorum
P. cyclostomus
P. ciliatus
P. insularis
P. multifasciatus
P. macronemus
P. barberinus
0.02
A. ensifera
A. semipunctata
0.02
(c)
S. punctatus
S. lividus
S. nigricans
S. limbatus
S. albifasciatus
S. gascoynei
S. fasciolatus
P. imparipennis
P. johnstonianus
P. phoenixensis
P. leucozonus
P. randalli
P. lacrymatus
S. pelicieri
P. lacrymatus
0.02
L. polynesica
L. dimidiatus
L. bicolor
L. dimidiatus
0.02
D. biocellatus
P. miles
D. zebra
P. antennata
P. radiata
0.02

Fig. 3. Major phylogeographic patterns observed among the 208 coral reef fish species with Indo-West Pacific range distribution and represented by at least 1 individual in each sample from the western Indian Ocean (WIO) (white), New Caledonia (NC) (grey) and French Polynesia (FP) (black). (a) Pattern I, where species are not spatially structured. From left to right: Kimura 2-parameter neighbor-joining (K2P-NJ) trees of *Sargocentron tiere*, *Uropterygius fuscoguttatus* and *Chromis viridis*. (b) Pattern II, where species display several allopatric mitochondrial lineages. From left to right: K2P-NJ trees of *Brotula multibarbata*, *Oxymonacanthus longisrostris* (here presented with nearest neighbor *Pervagor* spp.), *Pomacentrus pavo*, *Neoniphon samara*, *Parupeneus barberinus* and *Asterropteryx semipunctata*. (c) Pattern III, where species are either polyphyletic or paraphyletic and include mitochondrial lineage(s) of another species. From left to right: K2P-NJ trees of *Plectroglyphidodon lacrymatus* + *P. pelicieri*, *Labroides bicolor* + *L. dimidiatus* and *Pterois antennata* + *P. radiata* (nearest neighbor *Dendrochirus zebra*)

species level, as their DNA barcodes were nested within a monophyletic lineage encompassing a single nominal species. The present study is in agreement with previous large-scale studies in fishes that have shown a tight match between species identifications based on morphological characters and the boundaries of DNA barcode clusters (Ward et al. 2005, 2008, 2009, Hubert et al. 2008, 2011, 2012, April et al. 2011, Pereira et al. 2013).

The present study confirmed the occurrence of a substantial number of deeply divergent lineages among the 217 species with Indo-West Pacific distribution. Forty percent of the 185 species represented by >1 specimen per ocean consisted of at least 2 distinct mitochondrial lineages, one from the WIO and the other(s) from the western Pacific. In addition, nearly 10% of those 185 species were paraphyletic or polyphyletic. Only half of the species did not display distinct DNA barcode clusters between the Indian and the western Pacific sites. Of the latter, we found evidence of haplotype sharing in only 20 (21.5%), suggesting that the remaining 73 species (78.5%) exhibit haplotypes that are restricted to a single site, that is, phylogeographic structure without reciprocal monophyly. This result, however, is balanced by the small sample size per species in the present study and calls for a thorough assessment at larger spatial scales. If confirmed, this result may suggest that the high levels of cryptic diversity detected here are underestimated, as the more recently isolated lineages are likely to go undetected because of a lack of phylogenetic structure between oceans due to the persistence of shared ancestral polymorphism at the mitochondrial genome. As to the 12 species with several lineages that exhibited either paraphyletic or polyphyletic relationships, several cases may require morphological re-examination, as well as genetic analysis based on additional nuclear markers. These cases likely correspond to closely related lineages with conserved morphology and coloration patterns. These cases included *Myripristis* spp., *Sargocentron* spp., *Labroides dimidiatus* (Drew et al. 2008, Sims et al. 2014), *Pterois antennata*, *Kaupichthys diodontus* and *Dinematichthys iluocoeteoides*. Introgressive hybridization cannot be discarded as a potential explanation for the polyphyly and unexpectedly high intraspecific genetic distances observed in these species.

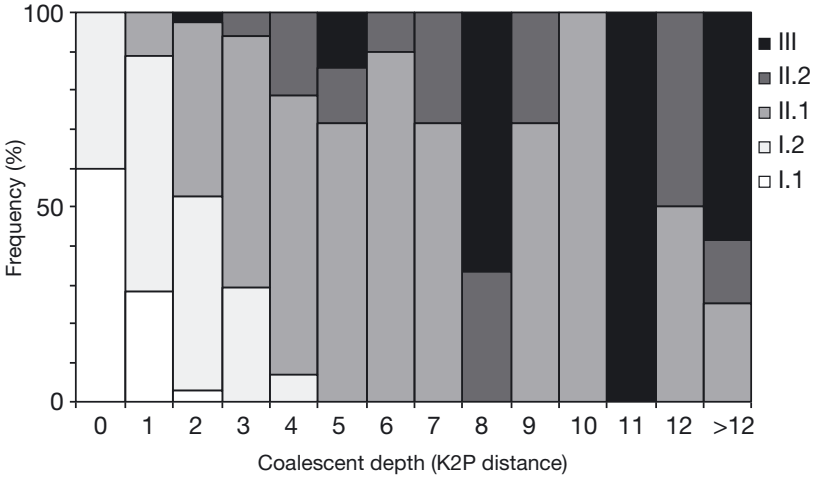

Fig. 4. Distribution of the relative frequencies of the phylogeographic patterns according to maximum intraspecific Kimura 2-parameter (K2P) distance (Table S3 in the Supplement at www.int-res.com/articles/suppl/m583p179_supp.xls). Pattern I.1, no phylogeographic structure among western Indian Ocean (WIO), New Caledonia (NC) and French Polynesia (FP) populations with haplotype sharing; pattern I.2, no phylogeographic structure among WIO, NC and FP populations without haplotype sharing; pattern II.1, detection of cryptic lineages alternatively sorted between the Indian and Pacific oceans (i.e. reciprocal monophyly of oceanic lineages); pattern II.2, detection of cryptic lineages alternatively sorted between the Indian and Pacific oceans but displaying geographic paraphyly (i.e. more than 2 lineages); pattern III, species are either paraphyletic or polyphyletic

Geographic isolation and population fragmentation

The differences between the local-scale and regional-scale distributions of intraspecific K2P distances suggest that the evolutionary history of the spe-

cies with Indo-Pacific or pan-Pacific distribution has been predominantly affected by geographic isolation, a result previously suggested by recent phylogenetic and biogeographic studies (Hubert et al. 2011, Quenouille et al. 2011, Pellissier et al. 2014, Bender et al. 2017). For instance, genetic divergence was higher between sites located in different oceans than between sites within the western Pacific Ocean. This result could be expected considering that 2 main, broadly concomitant processes are known to have promoted the isolation of the Indian and western Pacific oceans since the late Miocene. First, the mosaic of large islands and orogenic arches that constitute the Indo-Australian Archipelago have repeatedly merged during periods of low sea levels (Voris 2000, Hall 2012, 2013), creating physical barriers between the Indian and Pacific oceans (the Indo-Pacific Barrier [Gaither et al. 2011, Gaither & Rocha 2013]). Second, coral reef refugia have been more fragmented during the Pleistocene in the central Indian Ocean than in the Pacific Ocean. This mid-Indian Ocean barrier has hampered connectivity between the populations of the WIO and the eastern Indian Ocean (Pellissier et al. 2014).

The differences observed between the distributions of intraspecific K2P distances at the local scale (WIO, NC, FP) and those at the Indo-West Pacific scale are consistent with repeated isolation of the populations of the WIO caused by the Indo-Pacific and mid-Indian Ocean barriers through the Pleistocene. The large genetic distances observed between the cryptic lineages of the WIO and FP or NC suggest that the geographic isolation of the WIO site is older than that of the 2 sites sampled in the Pacific Ocean, or at least had experienced less faunistic exchange due to its geographic isolation. This hypothesis is also supported by the higher number of allopatric cryptic lineages observed between the WIO and the 2 sites in the Pacific Ocean than between the 2 sites in the Pacific Ocean. Considering that cryptic lineages generally consist of sibling lineages alternatively distributed in the WIO or FP and NC, our results are consistent with the hypothesis that cryptic diversity arose through allopatric speciation. This hypothesis is further supported by the high number of new species, as well as recently resurrected species in species pairs with allopatric Indian and Pacific distributions (Randall 2005, Craig & Hastings 2007, Randall & Rocha 2009, Allen & Erdmann 2012, Randall & DiBattista 2013, Borsa et al. 2014, Uiblein & Gouws 2014, Allen et al. 2015, Randall et al. 2015). The settlement of the Indo-Pacific and mid-Indian Ocean barriers contributed to isolate WIO from the Pacific and the eastern Indian oceans. Considering the absence of sites from the eastern Indian Ocean in our dataset, it is still unclear which of these 2 barriers contributed most in promoting the emergence of the cryptic lineages observed in the WIO and to what extent these lineages are restricted to the WIO in the Indian Ocean. A recent review of phylogeographic patterns in reef organisms of the Indian Ocean has revealed that WIO populations frequently display distinct phylogeographic groups compared to the eastern Indian Ocean (Borsa et al. 2016). The authors also highlighted that phylogeographic groups in the eastern Indian Ocean are frequently differentiated from those observed in the western Pacific. Thus, these 2 major phylogeographic breaks associated with the Indo-Pacific and mid-Indian Ocean barriers are likely to have concomitantly isolated the WIO and may account for the present pattern of cryptic diversity and genetic divergence.

Geographic isolation was also inferred between the 2 locations in the Pacific Ocean because local and regional distributions of maximum intraspecific distance were significantly different. The pattern of genetic divergence, however, differed from the pattern observed when comparing WIO with FP and NC. The distribution of mean intraspecific distances showed no difference between NC or FP and NC+FP, while the difference was significant between WIO or NC and WIO+NC. This may result from a more recent origin of the geographic isolation between NC and FP, as also suggested by the smaller number of species that exhibit cryptic lineages with alternative distribution ranges. Alternatively, dispersal may have contributed to maintain higher cohesion of the species pools in NC and FP than in WIO and Pacific Ocean sites owing to geographic proximity. Geographic isolation within the Pacific Ocean, however, may have promoted the emergence of cryptic diversity through allopatric speciation. The present findings strongly contrast with a recent review of reef fish DNA barcoding from the Caribbean (Victor et al. 2015). In Caribbean reef fishes, the smaller percentage of species exhibiting cryptic lineages is related to a much higher connectivity between the Caribbean islands and smaller spatial scales than those considered here for the Indian and Pacific oceans.

LHT and the dynamics of genetic divergence among populations

Additional clues about the ecological determinants of allopatric divergence were provided by the analy-

sis of the correlations between genetic divergence among cryptic lineages and LHT. Of the 3 LHTs examined here (i.e. SL, PLD, GRS), the one related to population connectivity (PLD) was not correlated with maximum intraspecific K2P distance in the species harboring 2 major mitochondrial lineages alternatively distributed in the western Indian vs. Pacific Ocean (pattern II.1). These results are consistent with the predominant impact of geographic isolation at large spatial scale, either western vs. eastern Indian Ocean or Indian vs. Pacific Ocean, where ecological connectivity has little impact (Mouillot et al. 2013). This result also suggests that PLD per se is not sufficient to explain connectivity patterns, as reported previously by authors who failed to detect any consistent relationship between PLD and population genetic structure (Bay et al. 2006, Riginos et al. 2011). Riginos et al. (2011) examined the population structure of dozens of benthic marine fish species for which they reported a preponderant impact of geographic isolation over dispersal ability. The lack of relationship between PLD and maximum intraspecific distance suggests that the impact of PLD on evolutionary dynamics is likely to depend on spatial scale and result from intricate interactions with other LHTs that may be predominant at larger spatial scales. Population divergence occurs at a larger spatial scale than the one involved in the ecological dynamics ruling connectivity, a trend that has been suggested in recent theoretical studies that failed to link population size and connectivity with rates of species diversification (Melián et al. 2010, Davies et al. 2011). The significant and negative correlation between GRS and maximum interspecific distance suggests, however, that a relationship of some sort exists between species' range and genetic divergence. This relationship might be a consequence of the influence of time on both GRS and population genetic divergence. A recent study reported that ancestral persistence in peripheral areas tends to produce old species with restricted range distribution at the margin of the Indian and Pacific oceans (Hodge & Bellwood 2015). This trend may be expected to contribute to a negative correlation between GRS and species age.

By contrast, a highly significant and negative correlation was detected between maximum intraspecific K2P distance and SL, meaning that larger species generally exhibit lower genetic divergence between cryptic lineages than smaller species. Body size is correlated with generation time (Martin & Palumbi 1993, Mooers & Harvey 1994, Smith & Donoghue 2008, April et al. 2013) and population size (Damuth 1981, Marquet et al. 1995), 2 factors that tightly interact with mutations to drive genetic drift (Kimura 1983) and result in varying substitution rates (Kingman 1982, Hudson 1983, Tajima 1983). The fact that the genetic divergence that accumulated among cryptic lineages in allopatry is negatively correlated with body size suggests that varying generation times and/or varying effective population sizes led to varying substitution rates. Disentangling the effect of generation time and effective population size requires genomic data beyond the scope of the present study; however, recent studies evidenced that effective population size has little influence on mitochondrial genetic diversity, including coral reef fishes (Bazin et al. 2006, Delrieu-Trottin et al. 2014). Alternatively, a relationship between species age and genetic divergence between cryptic lineages may account for this correlation. While old species may be expected to have more time to accumulate divergence among cryptic lineages than young species, there is no evidence in recent phylogenetic studies that body size and species age are correlated in reef fishes, as the distribution of species age largely overlaps among families with varying body size (Pellissier et al. 2014).

CONCLUSIONS

The DNA barcode reference library presented here constitutes the largest library assembled to date for Indo-West Pacific coral reef fishes, making it a useful reference for future studies. Present results have direct implications for the understanding of the evolutionary dynamics at play in coral reef fishes, particularly in a context of divergent interpretations about the origin of biodiversity patterns (Mora et al. 2003, Hubert et al. 2012, Gaither & Rocha 2013, Tornabene et al. 2015). Our results show that a substantial proportion of species with widespread distribution actually include several lineages that are likely to have originated through geographic isolation and diverged through the combined effects of mutation and genetic drift. This observation supports allopatric divergence as a major evolutionary engine in Indo-West Pacific coral reef fishes.

Acknowledgements. We are grateful to B. Bourgeois, G. Mou-Tham, C. Peignon and the participants of the RESICOD workshop in Noumea, December 2010, for their help during sample collection and to P. Irz and D. Ponton for help with photographing specimens. Fieldwork was conducted from the IRD research boats 'Coris' and 'Diodon', skippered by S. Tereua and N. Colombani, respectively. A. Amir and

C. Puech assisted with laboratory work. This work is part of the project 'Accurate SPEciEs Delimitation and IDentification of eukaryotic biodiversity using DNA markers' (@ SPEED-ID) proposed by F-BoL, the French Barcode of Life initiative. We are grateful to the BOLD staff and B. Victor for their help with the taxonomic updating of specimen records in BOLD. We are also grateful to the anonymous reviewers for insightful comments on earlier versions of this paper. This project was supported by the network Bibliothèque du Vivant funded by the CNRS, the MNHN, INRA and CEA (Genoscope) and by an MNHN grant from ATM Barcode. Sampling was in part supported by the Fondation pour la recherche sur la biodiversité (FRB, France) through the RESICOD project. All collecting and sequence data are available from BOLD in the projects IPCOM, SBF, MBFA, MBFC, FPFL and FPFLB in the Moorea Biocode—Fish container and RESIC in the Barcoding Fish (FishBOL) container. The sequences are also available from GenBank (see Table S2 in the Supplement at www.int-res.com/articles/suppl/m583p179_supp.xls for accession numbers). Supplementary data are available from the Mar Ecol Prog Ser website and from the Hal-IRD open access online repository (www.hal.ird.fr/; hal@ird.fr). P.B., A.D., N.H., M.K. and R.F.M. designed the project. P.B., M.K., R.F.M. and P.P. participated in the collection, identification and curation of samples. C.C. and A.D. supervised the laboratory work. P.B., N.H., M.K. and R.F.M. analyzed and interpreted the data. P.B. and N.H. wrote the manuscript. All authors have read and approved the final manuscript. This publication is contribution ISEM no. 2017-190 and IRD-UMR 250 no. 222.

LITERATURE CITED

Allen GR, Erdmann MV (2012) Reef fishes of the East Indies, Vols I–III. Tropical Reef Research, Perth

Allen GR, Erdmann MV, Kurniasih EM (2015) *Chrysiptera caesifrons*, a new species of damselfish (Pomacentridae) from the south-western Pacific Ocean. J Ocean Sci Found 15:16–32

April J, Hanner R, Mayden RL, Bernatchez L (2013) Metabolic rate and climatic fluctuations shape continental wide pattern of genetic divergence and biodiversity in fishes. PLOS ONE 8:e70296

April J, Mayden RL, Hanner RH, Bernatchez L (2011) Genetic calibration of species diversity among North America's freshwater fishes. Proc Natl Acad Sci USA 108: 10602–10607

Barber P, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. Proc Biol Sci 273:2053–2061

Bay LK, Crozier RH, Caley JM (2006) The relationship between population genetic structure and pelagic larval duration in coral reef fishes on the Great Barrier Reef. Mar Biol 149:1247–1256

Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. Science 312:570–572

Bender MG, Leprieur F, Mouillot D, Kulbicki M and others (2017) Isolation drives taxonomic and functional nestedness in tropical reef fish fauna. Ecography 40:425–435

Bergsten J, Bilton DT, Fujisawa T, Elliott M and others (2012) The effect of geographical scale of sampling on DNA barcoding. Syst Biol 61:851–869

Borsa P, Durand JD, Chen WJ, Hubert N, Muths D, Mou-Tham G, Kulbicki M (2016) Comparative phylogeography of the western Indian Ocean reef fauna. Acta Oecol 72:72–86

Borsa P, Hsiao DR, Carpenter KE, Chen WJ (2013) Cranial morphometrics and mitochondrial DNA sequences distinguish cryptic species of the longface emperor (*Lethrinus olivaceus*), an emblematic fish of the Indo-West Pacific coral reefs. C R Biol 336:505–514

Borsa P, Sembiring A, Fauvelot C, Chen WJ (2014) Resurrection of the Indian Ocean humbug damselfish *Dascyllus abudafur* (Forsskål) from synonymy with its Pacific Ocean sibling, *Dascyllus aruanus* (L.). C R Biol 337: 709–716

Briggs JC (1999) Extinction and replacement in the Indo-West Pacific Ocean. J Biogeogr 26:777–783

Briggs JC (2000) Centrifugal speciation and centres of origin. J Biogeogr 27:1183–1188

Briggs JC (2003) Marine centres of origin as evolutionary engines. J Biogeogr 30:1–18

Brodersen J, Seehausen O (2014) Why evolutionary biologists should get seriously involved in ecological monitoring and applied biodiversity assessment programs. Evol Appl 7:968–983

Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine metazoa. Annu Rev Mar Sci 3:471–508

Butcher BA, Smith MA, Sharkey MJ, Quicke DLJ (2012) A turbo-taxonomic study of Thai *Aleiodes* (*Aleiodes*) and *Aleiodes* (*Arcaleiodes*) (Hymenoptera: Braconidae: Rogadiniae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. Zootaxa 3457: 1–232

Craig MT, Hastings PA (2007) A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini. Ichthyol Res 54:1–17

Damuth J (1981) Population density and body size in mammals. Nature 290:699–700

Davies TJ, Allen AP, Borda-de-Água L, Regetz J, Melián CJ (2011) Neutral biodiversity theory can explain the imbalance of phylogenetic trees but not the tempo of their diversification. Evolution 65:1841–1850

Delrieu-Trottin E, Maynard J, Planes S (2014) Endemic and wide-spread coral reef fishes have similar mitochondrial genetic diversity. Proc R Soc B 281:20141068

Dettai A, Lautredou AC, Bonillo C, Goimbault E and others (2011) The actinopterygian diversity of the CEAMARC cruises: barcoding and molecular diversity as a multilevel tool for new findings. Deep-Sea Res II 58:250–263

Drew J, Allen GR, Kaufman L, Barber PH (2008) Endemism and regional color and genetic differences in five putatively cosmopolitan reef fishes. Conserv Biol 22: 965–975

Drew J, Barber PH (2009) Sequential cladogenesis of the reef fish *Pomacentrus moluccensis* (Pomacentridae) supports the peripheral origin of marine biodiversity in the Indo-Australian archipelago. Mol Phylogenet Evol 53: 335–339

Durand JD, Borsa P (2015) Mitochondrial phylogeny of grey mullets (Acanthopterygii: Mugilidae) suggests high proportion of cryptic species. C R Biol 338:266–277

Durand JD, Hubert N, Shen KN, Borsa P (2017) DNA barcoding grey mullets. Rev Fish Biol Fish 27:233–243

Froese R, Pauly D (eds) (2016) FishBase. www.fishbase.org (accessed on 1 June 2016)

Gaither MR, Bowen BW, Bordenave TR, Rocha LA and oth-

ers (2011) Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle. BMC Evol Biol 11:189

Gaither MR, Rocha LA (2013) Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. J Biogeogr 40: 1638–1648

Halas D, Winterbottom R (2009) A phylogenetic test of multiple proposals for the origins of the East Indies coral reef biota. J Biogeogr 36:1847–1860

Hall R (2012) Late Jurassic-Cenozoic reconstructions of the Indonesian region and the Indian Ocean. Tectonophysics 570-571:1–41

Hall R (2013) The palaeogeography of Sundaland and Wallacea since the late Jurassic. J Limnol 72:1–17

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci USA 101:14812–14817

Hodge J, Bellwood DR (2015) On the relationship between species age and geographical range in reef fishes: Are widespread species older than they seem? Glob Ecol Biogeogr 24:495–505

Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ

Hubert N, Delrieu-Trottin E, Irisson JO, Meyer C, Planes S (2010) Identifying early stages of coral reef fishes through DNA barcoding: a test case with the families Acanthuridae and Holocentridae. Mol Phylogenet Evol 55:1195–1203

Hubert N, Espiau B, Meyer C, Planes S (2015) Identifying the ichthyoplankton of a coral reef using DNA barcodes. Mol Ecol Resour 15:57–67

Hubert N, Hanner R (2015) DNA barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes 3:44–58

Hubert N, Hanner RH, Holm E, Mandrak NE and others (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLOS ONE 3:e2490

Hubert N, Meyer C, Bruggemann JH, Guérin F and others (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, Paradis E, Bruggemann JH, Planes S (2011) Community assembly and diversification in Indo-Pacific coral reef fishes. Ecol Evol 1:229–277

Hudson RR (1983) Testing the constant rate neutral allele model with protein sequence data. Evolution 37:203–217

Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PDN (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Phil Trans R Soc B 360:1835–1845

Jokiel P, Martinelli FJ (1992) The vortex model of coral reef biogeography. J Biogeogr 19:449–458

Kadarusman, Hubert N, Hadiaty RK, Sudarto, Paradis E, Pouyaud L (2012) Cryptic diversity in Indo-Australian rainbowfishes revealed by DNA barcoding: implications for conservation in a biodiversity hotspot candidate. PLOS ONE 7:e40627

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kingman JFC (1982) The coalescent. Stochastic Process Appl 13:245–248

Ko HL, Wang YT, Chiu TS, Lee MA and others (2013) Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLOS ONE 8: e53451

Kulbicki M, Parravicini V, Bellwood DR, Arias-Gonzalez JE and others (2013) Global biogeography of reef fishes: a hierarchical quantitative delineation of regions. PLOS ONE 8:e81847

Kuriiwa K, Hanzawa N, Yoshino T, Kimura S, Nishida M (2007) Phylogenetic relationships and natural hybridization in rabbitfishes (Teleostei: Siganidae) inferred from mitochondrial and nuclear DNA analyses. Mol Phylogenet Evol 45:69–80

Laboute P, Grandperrin R (2000) Poissons de Nouvelle-Calédonie. C. Ledru, Nouméa

Leprieur F, Albouy C, De Bortoli J, Cowman PF, Bellwood DR, Mouillot D (2012) Phylogenetic beta diversity: distinguishing between 'true' turnover of lineages and phylogenetic diversity gradients. PLOS ONE 7:e42760

Leray M, Agudelo N, Mills SC, Meyer C (2013) Effectiveness of annealing blocking primers versus restriction enzymes for characterization of generalist diets: unexpected prey revealed in the gut contents of two coral reef fish species. PLOS ONE 8:e58076

Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G (2010) Allopatric divergence and speciation in coral reef fish: the three-spot dascyllus, *Dascyllus trimaculatus*, species complex. Evolution 64:1218–1230

Luiz OJ, Allena AP, Robertson DR, Floeter SR and others (2013) Adult and larval traits as determinants of geographic range size among tropical reef fishes. Proc Natl Acad Sci USA 110:16498–16502

Marquet PA, Navarrete SA, Castilla JC (1995) Body size, population density, and the energetic equivalence rule. J Anim Ecol 64:325–332

Martin AP, Palumbi S (1993) Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci USA 90:4087–4091

Mat Jaafar TNA, Taylor MI, Mohd Nor SA, de Bruyn M, Carvalho GR (2012) DNA barcoding reveals cryptic diversity within commercially exploited Indo-Malay Carangidae (Teleosteii: Perciformes). PLOS ONE 7:e49623

Melián CJ, Alonso D, Vazquez DP, Regetz J, Allesina S (2010) Frequency-dependent selection predicts patterns of radiations and biodiversity. PLOS Comput Biol 6: e1000892

McCafferty S, Bermingham E, Quenouille B, Planes S, Hoelzer G, Asoh K (2002) Historical biogeography and molecular systematics of the Indo-Pacific genus *Dascyllus* (Teleostei: Pomacentridae). Mol Ecol 11:1377–1392

McPeek MA (2008) The ecological dynamics of clade diversification and community assembly. Am Nat 172: E270–E284

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLOS Biol 3:e422

Monaghan MT, Wild R, Elliot M, Fujisawa T and others (2009) Accelerated species inventories on Madagascar using coalescent-based models of species delineation. Syst Biol 58:298–311

Mooers AO, Harvey PH (1994) Metabolic rate, generation time, and the rate of molecular evolution in birds. Mol Phylogenet Evol 3:344–350

Mora C, Chittaro PM, Sale PF, Kritzer JP, Ludsin SA (2003) Patterns and processes in reef fish diversity. Nature 421: 933–936

Mouillot D, De Bortoli J, Leprieur F, Parravicini V, Kulbicki M, Bellwood DR (2013) The challenge of delineating biogeographical regions: nestedness matters for Indo-Pacific coral reef fishes. J Biogeogr 40:2228–2237

Myers RF (1999) Micronesian reef fishes: a comprehensive guide to the coral reef fishes of Micronesia, 3rd edn. Coral Graphics, Barrigada, Guam

Pante E, Puillandre N, Viricel A, Arnaud-Haond S and others (2015) Species are hypotheses: avoid connectivity assessments based on pillars of sand. Mol Ecol 24: 525–544

Parravicini V, Villéger S, McClanahan TR, Arias-González JE and others (2014) Global mismatch between species richness and vulnerability of reef fish assemblages. Ecol Lett 17:1101–1110

Pellissier L, Leprieur F, Parravicini V, Cowman PF and others (2014) Quaternary coral reef refugia preserved fish diversity. Science 344:1016–1019

Pereira LHG, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? BMC Genet 14:20

Plaisance L, Caley MJ, Brainard RE, Knowlton N (2011) The diversity of coral reefs: What are we missing? PLOS ONE 6:e25026

Quenouille B, Hubert N, Bermingham E, Planes S (2011) Speciation in tropical seas: allopatry followed by range change. Mol Phylogenet Evol 58:546–552

Randall JE (2005) Reef and shore fishes of the South Pacific: New Caledonia to Tahiti and the Pitcairn Islands. University of Hawai'i Press, Honolulu, HI

Randall JE, Allen GR, Steene RC (1997) Fishes of the Great Barrier Reef and Coral Sea. Crawford House Publishing, Bathurst, NSW

Randall JE, Connell AD, Victor BC (2015) Review of the labrid fishes of the Indo-Pacific genus *Pseudocoris*, with a description of two new species. J Ocean Sci Found 16: 1–55

Randall JE, DiBattista JD (2013) A new species of damselfish (Pomacentridae) from the Indian Ocean. Aqua Int J Ichthyol 19:1–16

Randall JE, Rocha LA (2009) *Halichoeres claudia* sp. nov., a new Indo-Pacific wrasse (Perciformes: Labridae), the fourth species of the *H. ornatissimus* complex. Zool Stud 48:709–718

Randall JE, Victor BC (2015) Descriptions of thirty-four new species of the fish genus *Pempheris* (Perciformes: Pempheridae), with a key to the species of the western Indian Ocean. J Ocean Sci Found 18:1–77

Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 7:355–364

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Riedel A, Sagata K, Suhardjono YR, Balke M (2013) Integrative taxonomy on the fast track—towards more sustainability in biodiversity research. Front Zool 10:15

Ricklefs RE (2003) A comment on Hubbell's zero-sum ecological drift model. Oikos 100:185–192

Riginos C, Douglas KE, Jin Y, Shanahan DF, Treml EA (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. Ecography 34:566–575

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145 : 1219–1228

Sims CA, Riginos C, Blomberg SP, Huelsken T, Drew J, Grutter AS (2014) Cleaning up the biogeography of *Labroides dimidiatus* using phylogenetics and morphometrics. Coral Reefs 33:223–233

Smith M A, Rodriguez JJ, Whitfield JB, Deans AR , Janzen DH, Hallwachs W, Hebert PDN (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc Natl Acad Sci U SA 105:12359–12364

Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc Natl Acad Sci USA 104:4967–4972

Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. Science 322: 86–89

Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. Genetics 105:437–460

Tänzler R, Sagata K, Surbakti S, Balke M, Riedel A (2012) DNA barcoding for community ecology—how to tackle a hyperdiverse, mostly undescribed Melanesian fauna. PLOS ONE 7:e28832

Tornabene L, Valdez S, Erdmann M, Pezold F (2015) Support for a 'center of origin' in the Coral Triangle: cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: *Eviota*). Mol Phylogenet Evol 82:200–210

Uiblein F, Gouws G (2014) A new goatfish species of the genus *Upeneus* (Mullidae) based on molecular and morphological screening and subsequent taxonomic analysis. Mar Biol Res 10:655–681

Victor BC (2015) How many coral reef fish species are there? Cryptic diversity and the new molecular taxonomy. In: Mora C (ed) Ecology of fishes on coral reefs. Cambridge University Press, Cambridge

Victor BC, Valdez-Moreno M, Vásquez-Yeomans L (2015) Status of DNA barcoding coverage for the tropical western Atlantic shore fishes and reef fishes. DNA Barcodes 3:85–93

Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J Biogeogr 27:1153–1167

Ward RD, Costa FO, Holmes BH, Steinke D (2008) DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa, but *Zeus faber* and *Lepidopus caudatus* each probably constitute two species. Aquat Biol 3:71–78

Ward RD, Hanner RH, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol 74: 329–356

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Phil Trans R Soc B 360:1847–1857

Winterbottom R, Hanner R, Burridge M, Zur M (2014) A cornucopia of cryptic species—a DNA barcode analysis of the gobiid genus *Trimma* (Percomorpha, Gobiiformes). ZooKeys 381:79–111

Winters KL, van Herwerden L, Choat JH, Robertson DR (2010) Phylogeography of the Indo-Pacific parrotfish *Scarus psittacus*: isolation generates distinctive peripheral populations in two oceans. Mar Biol 157:1679–1691

Appendix

BOLD TaxonID Tree

```
Title           : Merged Project
Date            : 6-March-2017
Data Type       : Nucleotide
Distance Model  : Kimura 2 Parameter
Marker          : COI-5P
Codon Positions : 1st, 2nd, 3rd
Labels          : Extra Info, Region, SampleID, Family, BIN uri
Filters         : Length > 200
Colorization    : [blue]=Stop Codons [red]=Contamination or misidentification

Sequence Count  : 3174
Species count   : 805
Genus count     : 299
Family count    : 84
Unidentified    : 3

BIN Count       : 885
```

Fig. A1. Summary data of midpoint-rooted neighbor-joining tree of the 3174 DNA barcodes collected from the total of 805 species analyzed in this study. The tree is available at www.int-res.com/articles/suppl/m583p179_app.pdf

Editorial responsibility: Per Palsbøll, Groningen, The Netherlands

Submitted: September 22, 2016; Accepted: August 23, 2017
Proofs received from author(s): November 10, 2017