Comparative phylogeography of the western Indian Ocean reef fauna

Philippe Borsa, Jean-Dominique Durand, Wei-Jen Chen, Nicolas Hubert, Delphine Muths, Gérard Mou-Tham, Michel Kulbicki

To cite this version:


HAL Id: ird-01278855
https://hal.ird.fr/ird-01278855
Submitted on 25 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Comparative phylogeography of the western Indian Ocean reef fauna

Philippe Borsa a,*, Jean-Dominique Durand b, Wei-Jen Chen c, Nicolas Hubert d, Delphine Muths e,f, Gérard Mou-Tham a, Michel Kulbicki a

a Institut de recherche pour le développement (IRD), UMR EMTROPI and LabEx “Corail”, Nouméa, New Caledonia
b IRD, UMR MARBEC, Montpellier, France
c Institute of Oceanography, National Taiwan University, Taipei, Taiwan
d IRD, UMR ISEM, Montpellier, France
e IFREMER, délégation de l’océan Indien, Le Port, La Réunion
f Université européenne de Bretagne, Rennes, France

* Corresponding author, present address. IRD-UMR 250 EMTROPI and LabEx “Corail”, c/o Universitas Udayana, Jl Raya Sesetan Gang Markisa no. 6, 80322 Denpasar, Indonesia.
E-mail addresses: philippe.borsa@ird.fr (P. Borsa), jean-dominique.durand@ird.fr (J.-D. Durand), wjchen.actinops@gmail.com (W.-J. Chen), nicolas.hubert@ird.fr (N. Hubert), muthsdelphine@yahoo.fr (D. Muths), gerard.mou-tham@ird.fr (G. Mou-Tham), michel.kulbicki@ird.fr (M. Kulbicki)
HIGHLIGHTS

- Checklists of reef fish species were used to assess provincial endemism in the Indian Ocean.
- The number of endemic species was highest in the western Indian Ocean.
- Phylogeographic patterns for 22 Indo-Pacific reef-associated species were reviewed and compared.
- Among-species discrepancies in phylogeographic patterns precluded generalization on the evolutionary processes at play.
- A proportion of likely cryptic species was highlighted.
ABSTRACT

Assessing patterns of connectivity at the community and population levels is relevant to marine resource management and conservation. The present study reviews this issue with a focus on the western Indian Ocean (WIO) biogeographic province. This part of the Indian Ocean holds more species than expected from current models of global reef fish species richness. In this study, checklists of reef fish species were examined to determine levels of endemism in each of 10 biogeographic provinces of the Indian Ocean. Results showed that the number of endemic species was higher in the WIO than in any other region of the Indian Ocean. Endemic species from the WIO on the average had a larger body size than elsewhere in the tropical Indian Ocean. This suggests an effect of peripheral speciation, as previously documented in the Hawaiian reef fish fauna, relative to other sites in the tropical western Pacific. To explore evolutionary dynamics of species across biogeographic provinces and infer mechanisms of speciation, we present and compare the results of phylogeographic surveys based on compilations of published and unpublished mitochondrial DNA sequences for 19 Indo-Pacific reef-associated fishes (rainbow grouper *Cephalopholis argus*, scrawled butterflyfish *Chaetodon meyeri*, bluespot mullet *Crenimugil* sp. A, humbug damselfish *Dascyllus abudafur/D. aruanus*, areolate grouper *Epinephelus areolatus*, blacktip grouper *E. fasciatus*, honeycomb grouper *E. merra*, bluespotted cornetfish *Fistularia commersonii*, cleaner wrasse *Labroides* sp. 1, longface emperor *Lethrinus* sp. A, bluestripe snapper *Lutjanus kasmira*, unicornfishes *Naso brevirosris*, *N. unicornis* and *N. vlamingii*, blue-spotted maskray *Neotrygon kuhlii*, largescale mullet *Planiliza macrolepis*, common parrotfish *Scarus psicattus*, crescent grunter *Terapon jarbua*, whitetip reef shark *Triaenodon obesus*) and three coastal Indo-West Pacific invertebrates (blue seastar *Linckia laevigata*, spiny lobster *Panulirus homarus*, small giant clam *Tridacna maxima*). Heterogeneous and often unbalanced sampling design, paucity of data in a number of cases, and among-species discrepancy in phylogeographic structure precluded any generalization regarding phylogeographic patterns. Nevertheless, the WIO might have been a source of haplotypes in some cases and it also harboured an endemic clade in at least one case. The present survey also highlighted likely cryptic species. This may eventually affect the accuracy of the current checklists of species, which form the basis of some of the recent advances in Indo-West Pacific marine ecology and biogeography.

Keywords: Indo-West Pacific; Biogeographic province; Endemism; Reef fish; Mitochondrial DNA; Parsimony network
1. Introduction

1.1. Background

Marine biologists define biogeographic provinces as regions of the ocean characterized by the presence of distinct communities that are thought to have some cohesion at the evolutionary timescale and found to be distinct from adjacent provinces (Spalding et al., 2007; Briggs and Bowen, 2012, and references therein). Thus-called marine biogeographic provinces generally harbour a proportion of endemic species. The extent of biogeographic provinces generally coincides with geologic and/or oceanographic boundaries (Spalding et al., 2007, and references therein; Briggs and Bowen, 2012; Obura, 2012; Veron et al., 2015). Quantitative delineation of biogeographic provinces based on 169 reef fish checklists from across all tropical oceans (Kulbicki et al., 2013a) showed less clear-cut province boundaries, than those described in the above surveys. The delimitation of biogeographic provinces is viewed as important to help establish patterns of connectivity at the community and the population levels, which in turn has relevance in resource management and conservation (Spalding et al., 2007; Obura, 2012; Kulbicki et al., 2013a; Veron et al., 2015).

Ocean currents transport larvae of sedentary marine species, allowing the connection of geographically-distant habitats suitable for adults. Larval transportation also enables sedentary marine species to colonize new habitats (e.g., recent volcanic islands), and to recolonize habitats where the species have been extirpated. Although the direction and intensity of currents may determine levels of contemporary gene flow, a species’ geographic structure over evolutionary time is more likely to be influenced by geographic distance between suitable habitats (Crandall et al., 2014). Albeit not a primary determinant of distribution range size, variability in pelagic larval duration may partly contribute to differences in the distribution of reef fish species in the Indo-Pacific, as species with longer larval duration have a greater chance to disperse over habitats unsuitable to adults than those with shorter pelagic larval duration (Scheltema, 1968; Woodland, 1990; Lester et al., 2007; Treml et al., 2012; Luiz et al., 2013). Adult body size is another factor affecting a species’ distribution range (Luiz et al., 2013). Larger adults have higher reproductive output and higher-quality offspring (Beldade et al., 2012; Hixon et al., 2014), which enhances the probability of achieving long-distance dispersal. At the intraspecific level, higher larval survival rate potentially translates into higher dispersal ability which in turn translates into potential for higher gene flow between populations. Population genetic differentiation decreases with pelagic larval duration in the tropical Indo-West Pacific fish family Siganidae (Lemer et al. 2007), where geographic range size also increases with pelagic larval duration (Woodland, 1990). Tropical Pacific fishes from families Chaetodontidae, Labridae and Pomacentridae show a similar correlation (Lester and Ruttenberg, 2005; Lester et al., 2007). Selkoe and Toonen (2011) found that pelagic larval duration and genetic differentiation estimates, i.e. S. Wright’s $F_{st}$ and the slope of the isolation-by-distance correlation (Rousset, 1997), typically reflect scales of dispersal if the sampling design is adequate. Gene flow counteracts genetic differentiation by natural selection and genetic drift (Slatkin, 1985), thus hampering adaptation to local conditions.

1.2. The western Indian Ocean (WIO) province
The present study focuses on the western Indian Ocean (WIO) province, the westernmost of the tropical Indo-West Pacific domain. The WIO province is defined here as the oceanic region bounded by the eastern coast of Africa from Somalia to Mozambique and extending to the East so as to include the Seychelles and Mascarene archipelagoes. This approximately coincides with Obura's (2012) and Spalding et al.'s (2007) definitions of the Western Indian Ocean province. Briggs and Bowen (2012) did not mention the WIO as a particular biogeographic province in terms of endemism or as a diversity hotspot.

The reef fish fauna of the WIO has attracted increased interest from the scientific community. Smith and Heemstra (1986) described a large proportion of the species occurring in South Africa and adjacent waters (Mozambique, Madagascar). Subsequently, a number of country-based checklists have been proposed in this region (e.g. Iles Eparses, the Comores archipelago, Seychelles, Mascarene Islands). These species checklists were used to delineate biogeographic regions (Kulbicki et al., 2013a), and to relate species richness to a number of large-scale environmental factors, including sea-surface temperature, island size and its degree of isolation, distance from the Coral Triangle, and reef area (Parravicini et al., 2013). These studies, along with previous work on reef fish macro-ecology (Santini and Winterbottom, 2002; Bellwood et al., 2012; Kulbicki et al., 2013b; Mouillot et al., 2013), suggest that the WIO has some specific characteristics. In particular, this part of the Indian Ocean holds more species than expected from modeling based on geomorphologic, biogeographic and environmental explanatory variables (Parravicini et al., 2013; Pellissier et al., 2014). This leads us to hypothesize that the WIO represents either a species sink, a center of origin, or a zone of overlap. Possible routes of migration for reef fishes include along the eastern African coast, from the Maldives-Chagos archipelagoes towards the Seychelles, or from temperate southern Africa towards the North-East. Obura (2012) (p. 1) posits that diversity patterns in WIO corals are consistent with the oceanography of the western Indian Ocean, “reflecting inflow of the South Equatorial Current, maintenance of high diversity in the northern Mozambique Channel, and export from this central region to the north and south, and to the Seychelles and Mascarene islands.” This amounts to suggesting that the WIO may harbour specific communities that both receive and export species.

1.3. Comparative phylogeography as an approach to test biogeographic hypotheses

Community assembly, the sum of the processes leading to the aggregation of species in ecological communities (Emerson and Gillespie 2008), results from intricate interactions at the regional and local scales (Ricklefs, 1987; Leibold et al., 2004; Leibold et al., 2010). Community assembly is affected by the dynamics of diversification, which drives the biodiversity build-up in a regional pool, and by the ecological dynamics within communities, which limits the number of species that coexist (Hubert et al., 2015). Historical biogeography has traditionally focused on the contribution of geological and paleoecological landscape histories on species diversification at the regional scale (Nelson and Platnick, 1981; Myers and Giller, 1988; Ricklefs and Schluter, 1993). For this, the distributions of species have been analyzed to detect coinciding geographical boundaries, which in turn have been linked to known past events (Cracraft and Prum, 1988; Morrone 1994; Morrone and Escalante, 2002). The introduction of genetic markers into historical biogeography
has led to the development of comparative phylogeography, which aims at uncovering shared patterns in the distribution of intraspecific genetic diversity (Avise, 2000). Phylogeography offers the tools and concepts to study the historical processes that may be responsible for the contemporary geographic distributions of gene genealogies. Thus, phylogeography addresses some of the conservation issues that are at the core of modern marine biogeography. This approach has proven to be informative in exploring evolutionary dynamics of species across biogeographical boundaries and in revealing the temporal dynamics underpinning the coincidence of species distributions (Bermingham and Moritz, 1998; Hubert et al., 2007; Lee and Johnson, 2009). Phylogeographic subdivisions generally corroborate biogeographic provinces based on species lists, with a notable exception at the Indian-Pacific boundary (Briggs and Bowen, 2012). Studying phylogeographic structure (i.e., generally at the intraspecific level) allows one to test hypotheses on geographic connectivity inferred from species endemicity and species richness data (e.g., Obura, 2012; Kulbicki et al., 2013a).

1.4. Objectives

The first objective of this study is to determine whether the reef fish assemblages found in the WIO hold specific properties by (1) comparing species richness in the WIO to other provinces in the Indian Ocean, and (2) examining if the general structure of the WIO’s reef fish assemblages is different from adjacent regions. Subsequently, one may hypothesize that the WIO is a species sink, or source, or zone of overlap, which may have important consequences in conservation.

Then, we review the phylogeography of coral reef species with wide Indo-West Pacific distribution to: (i) detect possible multiple-species patterns of differentiation for WIO populations; (ii) test the biogeographic hypotheses derived from data on reef fish assemblages. Our emphasis is on coral reef fishes, but comparative information on coral reef invertebrates is also included.

2. Materials and Methods

2.1. Partitioning the Indian Ocean into biogeographic provinces

Spalding et al. (2007) defined biogeographic regions or provinces based on expert advice and estimated that the resulting classification should apply to a large set of marine flora and fauna. Basing their work on coral species lists, Veron et al. (2015) did a cluster analysis which resulted in the eastern African coast, Madagascar, the Mascarene Islands, the Maldives, and Chagos being grouped together. However, the basic biogeographical units they used were derived from Spalding et al.’s (2007) provinces. Kulbicki et al. (2013a) distinguished biogeographical regions for reef fishes using a dissimilarity matrix to build clusters. These authors found that beyond large biogeographical units, the classification of reef fish into small biogeographic units was unstable. Nevertheless, their study showed that the eastern African reef fish fauna was distinct from that of Madagascar and the Mascarene Islands.

In the present work, provinces in the Indian Ocean were primarily defined according to Kulbicki et al. (2013a). For Australia, we chose the partition proposed by the Atlas of Living Australia (http://regions.ala.org.au/#rt=imcras). The Andaman-Nicobar archipelago, the Cocos
(Keeling) islands and Christmas Island were set apart as this group of islands has a depauperate fauna when compared to the nearby coastal areas of Sumatra and peninsular Malaysia. They are also thought to lie at the boundary between Indian and Pacific ocean faunas as they boast a number of Indian-Pacific geminate species pairs (Hobbs et al., 2009; Hobbs et al., 2014), a peculiarity that has not yet been documented for the coastal areas of Sumatra and peninsular Malaysia. We applied Ward's algorithm of aggregation (Legendre and Legendre, 2012) on reef fish checklists to help confirm biogeographical units defined from the literature (see above). Depending on the distance metric (Supplementary material, Fig. S1), Madagascar grouped either with eastern Africa (from tropical South Africa to Somalia), or with the Seychelles and the Mascarene Islands. Similarly, the Maldives and Chagos clustered together but depending on the distance metric would do so either with India, Sri Lanka and Lakshadweep, or with the Seychelles, Mascarene Islands, and other localities from eastern Africa (Supplementary material, Fig. S1). Therefore, we deemed it difficult to assign checklists to a given biogeographical region or province. The final classification proposed here (Fig. 1) combined the views of Spalding et al. (2007), Kulbicki et al. (2013a), Veron et al. (2015), and the results presented in Supplementary material, Fig. S1. This is intended as a consensus, reached with a certain part of subjectivity. The separation of the eastern African coast from the Madagascar-Seychelles-Mascarene group was a posteriori justified by the results of species-richness analysis.

2.2. Analysis of reef fish checklists

Checklists of reef fishes were obtained as detailed previously (Kulbicki et al., 2013a; Parravicini et al., 2013; Pellissier et al., 2014; Supplementary material, Table S1). Checklists of fishes from Australia were obtained from the Atlas of Living Australia website. Only fish species associated with either coral reefs or rocky reefs were retained for our analyses. For each species, size was obtained from FishBase (Froese and Pauly, 2013) and from other sources (see all references under label “General” in Supplementary material, Table S1). A geographic-range index was obtained for each species by counting the number of checklists on which it was found. A total of 289 species lists worldwide were used to build this index. Species lists were grouped into 10 Indian-Ocean provinces (Fig. 1). The geographic locations of checklist sites are given in Supplementary material, Table S1.

Species-richness analyses were restricted to reef fish checklists from the Indian-Ocean region. These checklists suggest that the WIO could be split into two sub-provinces, namely the eastern African (EAF) sub-province, and the Madagascar-Seychelles-Mascarene Islands (WII, for Western Indian Ocean Islands) sub-province. As confirmed by the present work, these sub-provinces show distinct patterns of reef fish endemism. Therefore, to explain some of the patterns we considered EAF and WII as separate biogeographic entities. Species were classed into four main categories of endemism according to their geographical range: (i) “local endemics”, i.e. species known from only one checklist; (ii) “province endemics”, i.e. species known from only two checklists within a province; (iii) “regional endemics”, i.e. species known from three to five checklists; (iv) “non endemics”, i.e. species found in six or more checklists. Species found on checklists from non-adjacent biogeographical provinces were not considered endemics.
2.3. Nucleotide-sequence data

We compiled mitochondrial nucleotide-sequence datasets for eight species or species complexes that matched the following criteria: (1) having a wide Indo-Pacific distribution; (2) being sampled in both the Indian and the Pacific Oceans; (3) having sampling localities in the Indian Ocean including, but not limited to, the WIO. The following species were selected: the humbug damselfish *Dascyllus abudafur* (Perciformes: Pomacentridae) and its Pacific-Ocean sibling *D. aruanus*, the areolate grouper *Epinephelus areolatus* (Perciformes: Serranidae), the blacktip grouper *E. fasciatus*, the honeycomb grouper *E. merra*, the bluespotted cornetfish *Fistularia commersonii*, the bluestripe snapper *Lutjanus kasmira* (Perciformes: Lutjanidae), the spiny lobster *Panulirus homarus* (Decapoda: Palinuridae), and the crescent grunter, *Terapon jarbua* (Perciformes: Teraponidae). The sequence datasets used for these species were gathered from previous projects of ours (Muths et al., 2012; Borsa et al., 2014; Liu et al., 2014; Muths et al., 2015), together with homologous sequences from the literature (Ward et al., 2005; Kawahara et al., 2008; Gaither et al., 2010; Asgharian et al., 2011; Lakra et al., 2011; Zhang, 2011; Hubert et al., 2012; Jeena, 2013; Zhuang et al., 2013; Alcantara and Yambot, 2014; Lavergne et al., 2014; Lavery et al., 2014; Schoelinck et al., 2014; Senevirathna and Munasinghe, 2014; Jackson et al., 2015), from the CRYOBANK barcode database (http://cryobank.sinica.edu.tw/) and from the GenBank nucleotide-sequence database (http://www.ncbi.nlm.nih.gov/). New cytochrome *b* gene sequences of 6 *D. aruanus* individuals from Madang, Bismarck Sea were produced according to the methods and protocols detailed in Liu et al. (2014). New *CO1* gene sequences were produced for *E. fasciatus* from Glorieuses Islands, Iles Eparses (*N* = 1), as well as *T. jarbua* from Iles Eparses (*N* = 7), Vietnam (*N* = 3), and Taiwan (*N* = 10). For this, genomic DNA was extracted using the PureLink® Genomic DNA Mini Kit (Life Technologies, Carlsbad CA) and a set of three primers (*FishF1*, *FishF2* and *FishR1*; Ward et al., 2005) was used to PCR-amplify a 650-bp fragment of the *CO1* gene. Purification of PCR products and sequencing were done by Macrogen (Amsterdam). CRYOBANK and GenBank accession numbers are provided in Appendix A. Sequences were edited using the 4PEAKS DNA sequence file viewer software (Mekentosj.com, Amsterdam; Griekspoor and Groothuis, 2005). For each species, sequences were aligned and trimmed to a common length under BioEdit ver. 7.1.11 (Hall, 1999).

2.4. Nucleotide-sequence data analysis

Median-joining parsimony networks were constructed using Network ver. 4.6.1.3 (Bandelt et al., 1999), based on the matrices of individual nucleotide sequences; the default settings of the program were selected. Nucleotide distances between haplogroups were estimated using MEGA6 (Tamura et al., 2013), based on the most likely nucleotide substitution model according to the Bayesian information criterion. The model chosen for the 1058-bp cytochrome *b* gene fragment in *D. abudafur* / *D. aruanus* was TN93+G (Tamura and Nei, 1993). Although the best model determined by MEGA6 for the 803-bp cytochrome *b* gene fragment in *E. merra* was HKY+G (Hasegawa et al., 1985), the model chosen was the second-best, TN93+G, which contrary to HKY+G, is available for estimating nucleotide distances in MEGA6. The model chosen for the 547-bp *CO1* gene fragment in *P. homarus* was T92+G (Tamura, 1992). The model chosen for the 475-bp
cytochrome $b$ gene fragment in *L. kasmira* and for the 551-bp *CO1* gene fragment in *T. jarbua* was K2P (Kimura, 1980).

In all networks, we distinguished haplotypes sampled in the Pacific Ocean from those sampled in the Indian Ocean. Within the latter, we distinguished those sampled in the WIO or from tropical South Africa [even though the latter formally belonged to the nearby southwestern Indian Ocean province (SWI) as defined in Section 2.1] from the rest of the Indian Ocean.

Levels of genetic divergence between populations were estimated with the $\phi_{ST}$ fixation index (Excoffier et al., 1992) computed using ARLEQUIN 3.5 (Excoffier and Lischer, 2010). For estimating $\phi_{ST}$, nucleotide differences were based on the nucleotide substitution model determined as the most likely according to the Bayesian information criterion. Because the HKY model is not implemented in ARLEQUIN, the more inclusive TN93 model was used instead. Significance of $\phi_{ST}$ for all possible pairwise population comparisons was assessed using 1,000 random permutations.

### 2.5. Survey of the phylogeographic literature

The core collection of the Web of Science ver. 2014 (Thomson Reuters, London; http://apps.webofknowledge.com/) was searched for additional articles on the marine phylogeography of the Indo-West Pacific. Combinations of the keywords “Indian”, “Pacific” and “phylogeography”, or their derivatives, were entered into the “topic” field of the Basic Search tool of the Web of Science. Among the articles thus retrieved, we retained those for which samples from both the Indian and the Pacific oceans were characterized genetically at the mitochondrial locus, including at least one sample from the WIO or tropical South Africa and one sample from the rest of the Indian Ocean. The eligible phylogeographic surveys in reef fishes and invertebrates included the peacock grouper *Cephalopholis argus* (Gaither et al., 2011), the scrawled butterflyfish *Chaetodon meyeri* (DiBattista et al., 2012), the cleaner wrasse *Labroides dimidiatus* (Drew et al., 2008; Sims et al., 2014), the longface emperor *Lethrinus* sp. A (one of the two biological species formerly under *L. olivaceus*; Borsa et al., 2013b), the blue sea star, *Linckia laevigata* (Crandall et al., 2014), the spotted and bluespine unicornfishes, *Naso brevirostris* and *N. unicornis* (Horne et al., 2008), the bignose unicornfish *N. vlamingii* (Klanten et al., 2007), the blue-spotted maskray *Neotrygon kuhlii* (Arlyza et al., 2013), the common parrotfish *Scarus psittacus* (Winters et al., 2010), the whitetip reef shark *Triaenodon obesus* (Whitney et al., 2012), and the small giant clam *Tridacna maxima* (Hui, 2012). An additional taxonomic survey of species in the family Mugilidae (Durand and Borsa, 2015) provided preliminary insights into the phylogeographies of the bluespot mullet, *Crenimugil* sp. A (one of the cryptic species under *Moolgarda seheli*; see Durand et al., 2012), and the largescale mullet, *Planiliza macrolepis* (formerly *Chelon macrolepis*; Durand et al., 2012). While grey mullets are generally not understood as coral reef-associated fishes, the habitats of *Crenimugil* sp. A and *P. macrolepis* include reef-associated mangroves and reef flats and lagoons.

The topology of the phylogenetic trees of *Labroides dimidiatus* mitochondria (Drew et al., 2008; Sims et al., 2014), together with correlated differences in morphology, imply that *L. dimidiatus* consists of two distinctly evolved entities which either have retained morphological and behavioural similarity, or have converged morphologically and behaviourally. It is thus sensible to treat the two forms of *L. dimidiatus* separately. One form, which corresponds to mitochondrial clade 1 of Sims et al. (2014), is distributed from the western Indian Ocean to the western Pacific Ocean;
the other form (clade 2) has so far been sampled exclusively in the southwestern Pacific (Drew et al., 2008; Sims et al., 2014). We examined the phylogeographic structure of the clade 1 form, coined *Labroides* sp. 1 hereafter.

### 3. Results

#### 3.1. Species richness patterns inferred from reef fish assemblages

The total number of reef fish species per province (Fig. 1A; Supplementary material, Table S2) varied from 612 in the subtropical southeastern Indian Ocean province (SCI), to 1931 in the adjacent Western Australian province (WAU). The WIO province (*N* = 1826) had nearly as many species as WAU but it was the largest in size of all Indian Ocean provinces. When we split the WIO into two provinces, namely WII and EAF, the WII had 1572 species and the EAF had 1532. The WIO had a number of reef fish species higher than, or equal to that in the East India–Thailand-Malaysia province (EI) (*N* = 1728), even though the latter is much closer to the Coral Triangle, the epicenter of Indo-West Pacific species richness for reef fishes (Briggs, 2003; Bellwood et al., 2005; Allen and Erdmann, 2012). Species richness in the WII sub-province was also higher than in the northwestern Indian Ocean (NWI) province. No gradient in species richness was observed among provinces.

The number of endemic species was much higher in the WIO than in any other region of the Indian Ocean (Fig. 1B; Supplementary material, Table S2). The WIO also had a high percentage of endemics (11.2%), close to those of the Red Sea/NWI province (13.4%) or the SWI province (13.3%). The majority of these endemics were regional endemics. No correlation was found between the number of endemics, or their proportion or their level of endemism, and sea-surface temperature, reef area, island size, degree of isolation, or distance to the Coral Triangle (as the assumed epicenter of Indo-West Pacific marine species richness). Within the WIO, endemic species were more abundant and more frequent in the WII than in the EAF. In the latter, the numbers and proportions of local and provincial endemics were among the lowest in the Indian Ocean.

Endemic reef fish species showed among-region variation in size distributions (Supplementary material, Fig. S2). The WIO displayed a heterogeneous composition, with the size distribution of endemic species in the EAF being rather similar to that in SWI. Both EAF and SWI had larger endemic species than the other regions of the Indian Ocean. All the other provinces were dominated by small endemic species (with a peak in size at 5-10 cm). However, the WII also showed a mode at 15-30 cm, like the continental African provinces, suggesting that similar processes may be at play in EAF, WII and South Africa.

#### 3.2. Multiple-species phylogeographic survey

In the pair *Dascyllus abudafur / D. aruanus*, mitochondrial haplotypes were split into two main haplogroups (Fig. 2) separated by 0.6% net nucleotide distance. One haplogroup comprised all haplotypes sampled in the Indian Ocean, the other one comprised all haplotypes sampled in the Pacific Ocean. Most haplotypes from the Red Sea, all private, formed a sub-lineage nested within the WIO haplogroup. This suggests that Red Sea haplotypes derive from either the WIO, or from a geographically larger population that encompasses the WIO. Population-pairwise *φ*<sub>ST</sub> estimates
(Supplementary material, Table S3) showed no heterogeneity within the WIO ($\phi_{ST}$ not significantly different from 0), while all comparisons between WIO populations and the other Indian-Ocean population, i.e. the Red Sea (RS) ($\phi_{ST} = 0.353-0.386$) were highly significant, indicating strong restriction in gene flow.

Strong phylogeographic structure was also evident in *Epinephelus areolatus* (Fig. 3). The only haplotype sampled in the WIO clustered with the other Indian-Ocean haplotypes (except those sampled in Western Australia) whereas all Pacific-Ocean haplotypes clustered together with Western Australian haplotypes as a separate haplogroup. Three distinct haplogroups were observed in *E. fasciatus* (Fig. 4A). The longest branch of the network separated haplotypes found exclusively in the WIO and the Red Sea, from the rest, i.e. from the central and eastern Indian Ocean and from the Pacific Ocean. No clear geographic segregation was observed within the latter ensemble, although haplotypes from different geographic origins were apparently not distributed randomly. Most haplotypes from the Coral Triangle clustered around the dominant haplotype, while the second-dominant haplotype mostly comprised haplotypes from the central and eastern Indian Ocean. The mitochondrial haplotype network produced for *E. merra* (Fig. 5A) can be easily superimposed on the map of samples (Fig. 5B), indicating phylogeographic structure. An Indian vs. Pacific phylogeographic partition was evident, where haplogroups from the two oceans were separated by 4.6% nucleotide distance, pointing to a geographic barrier between the two mitochondrial forms. The geographic barrier is located east of the Maldives and west of New Caledonia. Indian-Ocean haplotypes clustered into two distinct haplogroups, each centered around a dominant haplotype in a star-like fashion. One of the two haplogroups exclusively comprised haplotypes sampled from the WIO. The other haplogroup, which appeared to be an offshoot of the first haplogroup, comprised all haplotypes sampled in the Maldives together with a proportion of WIO haplotypes. In other words, the Maldives haplogroup was nested within the WIO haplogroup, suggesting past asymmetrical gene flow from the WIO, or from a geographically larger population encompassing the WIO, towards the central Indian Ocean. Population-pairwise $\phi_{ST}$ estimates (Supplementary material, Table S4) indicated substantial heterogeneity within the WIO (with $\phi_{ST}$ up to 0.344) and highly significant differences with the only other sample from the Indian Ocean, i.e. the Maldives (MAL) (average $\phi_{ST} = 0.269 \pm 0.125$).

No obvious phylogeographic structure was observed in *Fistularia commersonii*, although all haplotypes from Baja California clustered as a single branch of the network and the two haplotypes from Reunion Island (WIO) were distant from the rest of the network (Fig. 6).

*Lutjanus kasmira* haplotypes were organized as a star-like network with a dominant central haplotype that was shared between the WIO, the central and eastern Indian Ocean, and the Pacific Ocean. Little heterogeneity in haplotype frequencies was visible overall, except for the occurrence of a derived, predominantly Pacific-Ocean haplogroup (Fig. 7). Population-pairwise $\phi_{ST}$ estimates (Supplementary material, Table S5) singled out the Marquesas population ($\phi_{ST} = 0.499-0.627$) but otherwise showed only weak differentiation, with population-pairwise $\phi_{ST}$ values ranging from 0 to 0.077. The $\phi_{ST}$ estimates between the WIO and the other Indian-Ocean populations ranged from 0 to 0.050 and were hardly significantly different from 0 considering adjustment for multiple tests (Rice, 1989).

Two major haplogroups were observed in *Panulirus homarus*, separated by 5.8% net nucleotide distance at the *COI* locus (Fig. 8A). While the dominant haplogroup included haplotypes
from the whole distribution of *Panulirus homarus*, the second haplogroup (delineated by a dashed ellipse on Fig. 8A) comprised haplotypes from the WIO exclusively. The WIO haplogroup, which corresponds to morphological subspecies *P. h. rubellus* (Lavery et al., 2014), was discarded for $\phi_{ST}$ analysis. At the Indian-Ocean scale, the dominant haplogroup showed no obvious structure, while the $\phi_{ST}$ analysis distinguished the northwestern Indian-Ocean samples from all the rest ($\phi_{ST} = 0.108 \pm 0.106$) (Supplementary material, Table S6).

Three main haplogroups were observed in *Terapon jarbua* (Fig. 9). A haplogroup that included all haplotypes sampled in Juan de Nova, WIO, together with a haplotype from the Red Sea, was separated from the rest of the nucleotide sequence dataset by 7.6% net nucleotide distance at the *CO1* locus. Within the latter, a predominantly Indian haplogroup was separated from a predominantly Pacific haplogroup by 2.6% net nucleotide distance. High levels of genetic differentiation were estimated between populations, but no clear geographic pattern was apparent. The reason for this was the geographically widespread occurrence of the three haplogroups, in variable proportions. The Iles Éparses sample was strongly differentiated from the other populations ($\phi_{ST} = 0.472-0.628$; Supplementary material, Table S7) because all of its individuals harboured the Juan de Nova haplotype which was not sampled elsewhere except in the Red Sea.

### 3.3. Additional cases drawn from the phylogeographic literature

Fourteen additional cases were investigated (Supplementary material, Figs. S3-S14; Durand and Borsa, 2015), eleven of which concerned reef-associated bony fishes. In *Cephalopholis argus*, two major haplogroups were observed, with an Indian vs. Pacific phylogeographic partition. However, the 'Indian' haplogroup also harboured haplotypes sampled in the Pacific Ocean and vice versa, suggesting either incomplete lineage sorting, or secondary contact between the two oceans (Gaither et al., 2011). Haplotypes sampled in the WIO were nested within the 'Indian' haplogroup exclusively, suggesting that 'Pacific' haplotypes under Gaither et al.’s (2011) overlap hypothesis have not reached the WIO (Supplementary material, Fig. S3). In *Chaetodon meyeri* (Supplementary material, Fig. S4), only moderate Indian vs. Pacific phylogeographic partition was observed, suggesting weak differentiation between populations from the two oceans. In *Labroides* sp. 1, a loose Indian vs. Pacific genetic differentiation was visible (Supplementary material, Fig. S5), compatible with isolation by distance (Wright, 1942). This is the hypothesis favoured by Sims et al. (2014) on the basis of observed correlation between $\phi_{ST}$ and geographic distance. From the limited dataset available in *Lethrinus* sp. A (Supplementary material, Fig. S6), we observed a clear Indian vs. Pacific phylogeographic partition. In *Linckia laevigata* (Supplementary material, Fig. S7), two major clusters of haplotypes were observed, one of which included a few WIO haplotypes. This cluster included a substantially higher proportion of Indian-Ocean haplotypes than the other cluster, which was predominantly Pacific. The dominant WIO haplotypes occupied a central position in the network, implying that one, the other, or possibly both, of the two major haplogroups may derive from an ancestral WIO stock. No phylogeographic structure was observed in *Naso brevirostris, N. unicornis and N. vlamingii* (Supplementary material, Figs. S8-S10). No phylogeographic structure was visible from the haplotype network in *Scarus psittacus* (Supplementary material, Fig. S12) although peripheral genetic differentiation has been reported. Genetic differences in *Triaenodon obesus* indicated geographic isolation between populations either side of the Indo-Pacific barrier.
Strong phylogeographic structure was observed in *Tridacna maxima* across the Indo-Pacific barrier and within the Indian Ocean (Supplementary material, Fig. S14). Within the Indian Ocean, three distinct *T. maxima* clades were evident from maximum-likelihood analysis. One clade corresponded to Hui’s (2012) clade 5, exclusively found in the WIO; a second clade corresponded to Hui’s (2012) clade 2, exclusive to the eastern Indian Ocean and the Sunda Shelf; a third clade corresponded to Hui’s (2012) clades 4 and 6 from the WIO and the Red Sea (Supplementary material, Fig. S14). Similarly strong, parapatric geographic structure was observed in *Neotrygon kuhlii*, where 7 clades had been sampled by Arlyza et al. (2013), all from the sole Coral Triangle region (Supplementary material, Fig. S11). The two *Neotrygon kuhlii* haplotypes sampled from the WIO formed a lineage distinct from all other sites sampled in the Indian Ocean and on the Indian-Ocean side of the Coral Triangle.

In addition, little genetic differentiation was evident in *Crenimugil* sp. A, where the same mitochondrial haplotype was sampled from the WIO to the southwestern Pacific. However, a distinct sub-clade from the northwestern Indian Ocean was observed (Durand and Borsa, 2015). In *Planiliza macrolepis*, haplotypes sampled in the Maldives and further east belonged to a clade distinct from haplotypes sampled in the western and northwestern Indian Ocean (see figures 1A and 4E of Durand and Borsa, 2015).

### 4. Discussion

#### 4.1. Species diversity and endemism patterns

The WIO showed higher reef-fish species richness than provinces geographically closer to the Coral Triangle, the epicenter of species richness in the Indo-West Pacific, for reef fishes (Briggs, 2003; Bellwood et al. 2005; Allen and Erdmann, 2012) and more generally for the coastal marine fauna and flora (Sanciangco et al., 2013). This was unexpected for several reasons. First, the number of species usually decreases as distance from the biodiversity center increases (Briggs, 2003; Parravicini et al., 2013). Pellissier et al. (2014) proposed that this is in part due to evolutionary history, with regions close to the assumed diversity center also being close to coral refuges during cold climate episodes. In addition, the WII has many small islands which typically harbour low species richness, because species richness decreases with island size (Parravicini et al., 2013). One would therefore expect fewer species in the WII than on the continental shelf of Africa (EAF and SWI). We observed the opposite.

The number of endemic species in the WIO was higher than in the northwestern part of the Indian Ocean (NWI province), as already shown by Kulbicki et al. (2013a), despite the fact that NWI has the highest rate of local endemism in the Indo-Pacific after Easter Island and Hawaii (Allen, 2008). We do not know if these endemic species are of local origin or if they are relicts from wider, past distributions. Dedicated phylogeographic surveys of these endemic species, that would also examine phylogenetically related species would help uncover their origin(s). The analysis of body-length distributions nevertheless provided some interesting information. Along the western shores of the Indian Ocean (NWI, EAF, SWI), endemic species were larger than elsewhere in the Indian Ocean. This is to be related to a similar situation in Hawaii where local endemics are larger than endemics in other parts of the Pacific Ocean (DeMartini and Friedlander, 2004). Hawaii, just as the
eastern African coast, lies at the edge of its realm. Species size is a primary determinant of species
distribution and large species tend to be better dispersers and better colonizers than small species
(Luiz et al., 2013). As a consequence, larger species are more frequent in assemblages on small or
remote islands, just as in the present case off the eastern African coast. In the other part of the WIO,
WII endemics are, in majority, small species, although with a second mode of larger species. This
suggests a complex origin for endemics. Small-size endemics dominate in the northwestern part of
the Indian Ocean, which suggests that the continental or island coastal type was not a major
determinant of the distribution of sizes. Both the Red Sea and Madagascar held coral reef refuges
during the last 3 million years (Pellissier et al., 2014). These refuges could have sheltered small
species which subsequently may have not re-colonized a wide range and therefore became
demic.

4.2. Variability in phylogeographic structure

As shown by the case studies used for the present review, geographic sampling design in
phylogeographic surveys of Indo-West Pacific reef species was mostly uneven and unbalanced.
Even the aggregative survey of Linckia laevigata by Crandall et al. (2014), presented (p. 399) as
"one of the most geographically comprehensive genetic studies of any Indo-Pacific species to date",
had gaps in its sampling design, with the western and central Indian-Ocean regions being
noticeably under-sampled. This, together with the low number of comprehensive case studies
available, are weaknesses of the present multiple-species phylogeographic survey.

Remarkable variability in phylogeographic patterns and in intensity of genetic differentiation
was observed. In several cases (Dascyllus abudafur/D. aruanus, Epinephelus areolatus, E. merra,
Lethrinus sp. A, Tridacna maxima), a marked Indian vs. Pacific partition was observed, suggesting
geminate species formed by allopatry on either side of the Indo-Pacific barrier. Weaker Indian vs.
Pacific phylogeographic partition was apparent in Cephalopholis argus, Labroides sp. 1 and Linckia
laevigata, which could result either from secondary contact and incomplete re-homogenization
across the Indo-Pacific barrier (C. argus; Gaither et al., 2011), incomplete lineage sorting, or
isolation by distance. In all three cases, as in a number of other cases from the phylogeographic
literature reviewed by Carpenter et al. (2011) and Gaither and Rocha (2013), the phylogeographic
structure was consistent with incipient genetic divergence of Indian vs. Pacific populations. Among
reef-associated animals, this also includes the neritid snail Nerita albicilla (Crandall et al. 2008) and
the bigscale soldierfish Myripristis berndti (Muths et al., 2011). Another phylogeographic break
located in the mid-Indian Ocean was observed in both E. fasciatus and Planiliza macrolepis. In these
and several of the above cases, the western Indian Ocean populations showed phylogeographic
originality compared to the eastern Indian Ocean populations, presumably because of the
geographic-isolation effect of the mid-Indian Ocean oceanic barrier west of the Lakshadweep-
Maldives-Chagos archipelagoes.

Little phylogeographic structure was evident in Fistularia commersonii, where the dominant
haplotypes were present in all regions sampled. No phylogeographic structure was detected in
seven other species, namely Chaetodon meyeri, Crenimugil sp. A, Lutjanus kasmira, Naso brevirostris,
N. unicornis, N. vlamingii, and Panulirus homarus (P. h. rubellus and the Marquesas sub-lineage
excluded). Among reef fishes, Chaetodon ornatissimus (the sister-species of C. meyeri; DiBattista et
al., 2012), *F. commersonii*, *L. kasmira* and *N. lituratus* (congeneric with *N. brevirostris* and *N. vlamingii*, and sister species to *N. unicornis*; Klanten et al., 2004) have pelagic larval durations >30 d, i.e. longer than the other reef fishes considered in this review (see Supplementary material, Table S8). The *Dascyllus abudafur/D. aruanus* species pair, which had the highest levels of population differentiation, is also among those with the shortest pelagic larval duration (18.6-24.3 d). However, this handful of case studies is not sufficient to draw conclusions on a possible effect of pelagic larval duration on phylogeographic structure in Indian-Ocean reef fishes. Complex phylogeographic structure was uncovered in *Terapon jarbua*, where the WIO, the northwestern Indian Ocean, and the Coral Triangle might be centers of diversification. One may suspect that *T. jarbua* actually consists of two, perhaps three biological species whose geographic distributions partly overlap. This possibility cautions against interpreting $\phi_{ST}$ estimates without prior sorting of samples by species.

The relative lack of consistency in the phylogeographic structures apparent from the 22 cases that we examined precludes any generalization about the evolutionary processes that have shaped the phylogeographic structure of species in the WIO. More research is certainly necessary to possibly draw generalities, and possibly alter this conclusion. Nevertheless, novel information arose from the present review. For instance, the observed phylogeographic structures in some species (*D. abudafur*, *E. merra*, *Linckia laevigata*) tended to designate the WIO as a possible source of haplotypes. The WIO also harboured endemic clades in *Panulirus homarus* (*P. h. rubellus*) and in *Tridacna maxima*, suggesting either long-term past geographic isolation or unique ecological characteristics suitable to the latter.

### 4.3. Need for taxonomic updates

Previous phylogeographic work has shown that several widespread Indo-West Pacific reef species actually consist of distinct evolutionary units that qualify as separate species (e.g. Wörheide et al., 2008; Fauvelot and Borsa, 2011; Vogler et al., 2012; Borsa et al. 2013a; Hoareau et al., 2013;Huelsken et al., 2013). This is also the case of *D. aruanus* which is now recognized to consist of the geminate Indo-Pacific species pair *D. abudafur/D. aruanus* (Borsa et al., 2014), and of *Labroides dimidiatus* (Sims et al., 2014; see section 2.4), whose pattern of cryptic speciation parallels that of *Lethrinus olivaceus* (Borsa et al., 2013b). *L. olivaceus* consists of two biological species provisionally designated as *Lethrinus* sp. A in the Indo-West Pacific, and *Lethrinus* sp. B, from the western and central Pacific (Borsa et al., 2013b). The subspecies of *P. homarus* endemic to the WIO, *P. h. rubellus*, should similarly be raised to the species status, because of concurring morphological and genetic distinctness (Lavery et al., 2014). Hypothesizing that the distinction of the *rubellus* form pertains to intraspecific variation does not hold given the long pelagic larval stage in *P. homarus* which lasts up to several months, as in other *Panulirus* species (Goldstein et al., 2008; Phillips and Matsuda, 2011). It has also been argued that the western and northwestern-Indian Ocean clade of *Planiliza macrolepis* represents a species distinct from the populations east of the mid-Indian Ocean barrier, for which the provisional species name *Planiliza* sp. H has been proposed (Durand and Borsa, 2015). The present study highlights additional cases where the current taxonomy of species may not reflect their evolutionary history, thus flagging potential cryptic species. For instance, *E. areolatus*, *E. merra* and *Lethrinus* sp. A may each consist of a pair of geminate Indian / Pacific
species. Phylogeographic patterns in *E. fasciatus* (present work) lead us to similarly hypothesize two cryptic species separated by the mid-Indian Ocean barrier. One of the potential biological species under *Terapon jarbua* appears to be endemic to the western and northwestern Indian Ocean. Taxonomic investigations of *Neotrygon kuhlii* and *Tridacna maxima* are also warranted.

Based on preliminary barcoding surveys, the occurrence of cryptic species might well occur in up to 30%-60% of Indo-West-Pacific reef fishes (Zemlak et al., 2009; Hubert et al., 2012). If this is confirmed by further taxonomic investigations, the species checklists at the basis of a number of recent advances in the biogeography and ecology of reef fishes (e.g. Bellwood et al., 2005; Bellwood et al., 2012; Kulbicki et al., 2013a; Mouillot et al., 2013; Par ravicina et al., 2013; Pelissier et al., 2014) will likely become partly obsolete. Consequently, the delineation of biogeographic provinces evoked in the present paper might also have to be revised. However, if most of the additional cryptic species result from thus-far neglected Indian vs. Pacific differences, it is unlikely that the partition into biogeographic provinces within each ocean will change drastically from the current one.

**Authors’ contributions**

Designed the study: PB, WJC, JDD, MK. Contributed samples / materials / analysis tools: PB, WJC, JDD, MK, GMT, DM. Analyzed and interpreted the data: PB, JDD, MK, DM. Wrote the paper: PB, NH, MK. All authors read and approved the final manuscript.

**Acknowledgements**

This research was supported by France's "Iles Eparses" program (2011-2013) managed by CNRS-Institut écologie et environnement (InEE) with the financial support of CNRS-InEE, CNRS-Institut national des sciences de l’univers (INSU), Institut de recherche pour le développement (IRD) and Agence des aires marines protégées (AAMP), and with the logistic support of Terres australes et antarctiques françaises (TAAF). It is the product of project PHYLIP on the comparative phylogeography of the benthic reef fauna of Iles Eparses in an Indo-Pacific context. We thank R. Beldade, G. Bernardi, J. Bourjea, K.E. Carpenter, P. Chabanet, T. Changeux, P. Courtès, P. Durville, T. Hoareau, C. Marteau, J. O’Donnell, D. Ponton, J.-F. Silvain, E. Teissier and M. Troussellier for stimulating discussions at various stages of this project. We are grateful to Mark A. Phuong, who kindly accepted to copy-edit the manuscript prior to final submission. We thank G.N. Mahardika (IBRC, Denpasar) for discussions on the taxonomy of *Panulirus homarus* and D. Ponton (IRD, Reunion) for preliminary analyses of morphological variation in *Terapon jarbua*. We are grateful to three anonymous reviewers for insightful comments, helpful suggestions, and some editing. Background maps of the Indo-West Pacific were edited from images downloaded from Digital Vector Maps, San Diego (http://digital-vector-maps.com/). Fish sampling at Iles Eparses was authorized under arrêté no. 2011-32 of TAAF and was done during a cruise of RV *Marion Dufresne*, 01-26 April 2011. This cruise offered to four of the co-authors (PB, WJC, JDD, GMT) an opportunity to meet and socialize, and to meet DM in Reunion Island. WJC acknowledges grant support from the Ministry of Science and Technology in Taiwan (MOST 101-2611-M-002-016-MY3). The research done on *Epinephelus merra* and *Lutjanus kasmira* was funded by the European Cooperation Fund
Appendix A. Database accession numbers for nucleotide sequences

CROYBANK accession number for E. fasciatus was ASIZP0800173. GENBANK accession numbers for D. abudafur / D. aruanus were KF754733-KF754795 and 6 additional cytochrome b gene sequences from Madang (GENBANK nos. KT258998-KT258994); those for E. areolatus were DQ107866-DQ107870, GU324187, FJ237756-FJ237763, HQ149838, HQ913572, HQ945841, JN208569-JN208572, JX674967-JX674969, KC593374, KC970469, KF009590-KF009591, KJ130964-KJ130965, KJ202151-KJ202152, KJ607969, KM077914, KP058460 and NC_020785; those for E. fasciatus were DQ107873-DQ107876, EU392207-EU392208, EU600145-EU600146, FJ459561-FJ459563, GU805082, JQ431717, JX093907, JX674994-JX674996, KC970470, KF009593, KF489582, KF714942, KF802940, KJ130967-KJ130969, KJ202153-KJ202155 and KJ594972-KJ594972, and an additional CO1 gene sequences from Glorieuses Islands (GENBANK no. KT258995); those for E. merra were AY78644, JN255254–JN255345 and JN545057–JN545096; those for F. commersonii were EF607383- EF607384, JF493489, JQ349978- JQ349979, JQ431740-JQ431741, KP053133-KP053237, KP260463 and NC_010274; those for L. kasmira were JF514414-JF514500 and FJ754049-FJ754133; those for P. homarus were JN418937, JN542716, JN591360, JQ229883-JQ229888, JQ229910-JQ229926, KC959890-KC959891, KF548571, KF715528-KF715550, KF715552, and KJ802748-KJ802782; those for T. jarbua included EU871691, EU871692, FJ237549, FJ265859, HQ149959-HQ149961, JF494663-JF494666, JN021254, JN021255, JQ342095-JQ342111, JX260979, KC241987, KC417308, KC774674, KC970423, KF009671, KF268188, KF268189, KJ466137 and KJ466138, and 20 additional CO1 gene sequences from Iles Eparses (GENBANK nos. KT231909-KT231915), Vietnam (GENBANK nos KT231926-KT231928) and Taiwan (GenBank nos KT231916-KT231925). Sequences nos. EU541344, HM909801 and KJ130970 labeled “Epinephelus fasciatus” were discarded because they were found outside the E. fasciatus CO1 haplotype cluster in a preliminary Neighbour-Joining tree of serranid barcodes, hence were thought to be possibly from misidentified specimens.

Appendix B. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.actao.2015.xxx

References


Perth, 1292 p.


Griekspoor, A., Groothuis, T., 2005. 4PEAKS: a program that helps molecular biologists to visualize and edit their DNA sequence files (http://mekentosj.com/4peaks/).


CAPTIONS TO FIGURES

Fig. 1. Reef fish species richness and endemism in the Indian Ocean. Abbreviations for provinces as following: AND Andaman-Cocos (Keeling)–Christmas; EAF Eastern Africa; EI Eastern India-Thailand-Malaysia; LMC Lakshadweep–Maldives–Chagos; NAU Northern Australia; NWI Northwestern Indian Ocean; SCI Spencer Gulf; SWI Southwestern Indian Ocean; WAU Western Australia; WII Western Indian Ocean Islands. A. Delimitation of biogeographic provinces with total number of reef fish species per province. Circle area proportional to number of species. B. Number of species with limited geographic range in each biogeographic province. Circle area proportional to total number of species with limited geographic range. Inset: categories of endemism according to number of checklists where a species occurs (see Materials and Methods). Numbers, from top to bottom: local, provincial, and regional endemics.

Fig. 2. Phylogeographic structure of humbug damselfish Dascyllus abudafur and D. aruanus. Includes cytochrome b gene sequence data from Liu et al. (2014). A. Median Joining parsimony network (Bandelt et al., 1999) of the partial cytochrome b gene sequences of 268 individuals from various localities in the Indo-West Pacific, including the WIO (black), the rest of the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences trimmed to 949 bp, starting at nucleotide site homologous to site no. 14475 of the mitochondrial DNA (= no. 46 of the cytochrome b gene) of Abudefduf vaigiensis (GENBANK no. NC_009064). Scale bar: one mutational step. B. Sampling locations.

Fig. 3. Phylogeographic structure of areolate grouper Epinephelus areolatus. Includes CO1 gene sequence data from Ward et al. (2005), Asgharian et al. (2011), He et al. (2013), Zhuang et al. (2013), Alcantara and Yambot (2014), Schoelinck et al. (2014), and GENBANK. A. Median Joining parsimony network (Bandelt et al. 1999) of the CO1 gene sequences of 37 individuals from various localities in the Indo-West Pacific, including the WIO (black), the rest of the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences trimmed to 547 bp, starting at nucleotide site no. 111 of the CO1 gene in this species (GENBANK no. NC_020785). Scale bar: one mutational step; number: number of mutations separating the three main haplogroups; cross: sampling site unknown (Nicô et al., 2012). Haplotypes sampled in Western Australia are distinguished by symbol Wa. B. Known sampling locations. Wa Western Australia.

Fig. 4. Phylogeographic structure of blacktip grouper Epinephelus fasciatus. Includes CO1 gene sequence data from Ward et al. (2005), Lakra et al. (2011), Hubert et al. (2012), Alcantara and Yambot (2014), Schoelinck et al. (2014), CRYOBANK, and GENBANK. A. Median Joining parsimony network (Bandelt et al. 1999) of the CO1 gene sequences of 31 individuals from various localities in the Indo-West Pacific, including the WIO (black), the rest of the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences trimmed to 508 bp, starting at nucleotide site homologous to site no. 133 of the CO1 gene of E. fuscoguttatus (GENBANK no. HQ174849). Scale bar: one mutational step; numbers: numbers of mutations separating the three main haplogroups. The single haplotype sampled in the Red Sea is distinguished by symbol Rs. B. Sampling locations. Rs Red Sea.
Fig. 5. Phylogeographic structure of honeycomb grouper *Epinephelus merra*. Includes cytochrome \( b \) gene sequence data from Muths et al. (2015), and GenBank. **A.** Median Joining parsimony network (Bandelt et al. 1999) of the cytochrome \( b \) gene sequences of 487 individuals from various localities in the Indo-West Pacific, including the WIO (black), the rest of the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences trimmed to 803 bp, starting at nucleotide site homologous to site no. 61 of the cytochrome \( b \) gene of *E. costae* (GenBank no. DQ197951). **Scale bar:** one mutational step; **number** indicates the number of mutations separating the two main haplogroups. **B.** Sampling locations.

Fig. 6. Phylogeographic structure of bluspotted cornetfish *Fistularia commersonii*. Includes CO1 gene sequence data from Zhang (2011), Hubert et al. (2012), Jackson et al. (2015), and GenBank. **A.** Median Joining parsimony network (Bandelt et al. 1999) of the CO1 gene sequences of 113 individuals from various localities in the Indo-West Pacific, including the WIO (black), the rest of the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences trimmed to 474 bp, starting at nucleotide site no. 201 of the gene (GenBank no. NC_010274; Kawahara et al., 2008). **Cross** indicates additional sequence of Kawahara et al. (2008), of unknown origin. **Scale bar:** one mutational step. **Re, Bc** haplotypes sampled in, respectively, Reunion Island and Baja California. **B.** Sampling locations, excluding La Paz, Baja California. **Re** Reunion Island.

Fig. 7. Phylogeographic structure of bluestripe snapper *Lutjanus kasmira*. Sequence data from Muths et al. (2012). **A.** Median Joining parsimony network (Bandelt et al., 1999) of the cytochrome \( b \) gene sequences of samples (total \( N = 935 \)) from various localities in the Indo-West Pacific, including the WIO (black), the rest of the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences trimmed to 475 bp, starting at nucleotide site homologous to site no. 14667 of the complete mitochondrial DNA of the species (= no. 274 of the cytochrome \( b \) gene) (GenBank no. NC_011578). **Scale bar:** one mutational step. **B.** Sampling locations.

Fig. 8. Phylogeographic structure of spiny lobster *Panulirus homarus*. Includes CO1 gene sequence data from Jeena (2013), Lavery et al. (2014), and Senevirathna and Munasinghe (2014); other data retrieved from GenBank. **A.** Median Joining parsimony network (Bandelt et al., 1999) of the partial CO1 gene sequences of samples from various localities in the Indo-West Pacific, including the WIO (black), other localities in the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences (total \( N = 98 \)) were trimmed to a common length of 547 bp, starting at nucleotide site homologous to site no. 135 of the mitochondrial DNA (= no. 135 of the CO1 gene) of *P. ornatus* (GenBank no. NC_014854). **Scale bar:** one mutational step; **numbers** indicates the number of mutations separating the two main haplogroups. **B.** Sampling locations.

Fig. 9. Phylogeographic structure of crescent grunter *Terapon jarbua*. Includes CO1 gene sequence data from Asgharian et al. (2011), Lakra et al. (2011), Lavergne et al. (2014), and GenBank. **A.** Median Joining parsimony network (Bandelt et al., 1999) of the CO1 gene sequences of samples from various localities in the Indo-West Pacific, including the WIO (black), other localities in the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences trimmed to 551 bp,
starting at nucleotide site homologous to site no. 5591 of the mitochondrial DNA (= no. 109 of the CO1 gene) of *Crenimugil crenilabis* (GenBank no. NC_017884). *Scale bar:* one mutational step; *numbers* indicate the numbers of mutations separating the three main haplogroups. Haplotypes sampled in the Red Sea and Juan de Nova are distinguished by symbols *Rs* and *Jn*, respectively. **B.** Sampling locations. *Rs* Red Sea; *Jn* Juan de Nova.
Figure 1 Borsa et al
Figure 2 Borsa et al
Figure 3 Borsa et al
Figure 4 Borsa et al
Figure 9 Borsa et al
Supplementary material to:

Comparative phylogeography of the western Indian Ocean reef fauna
P. Borsa, J.-D. Durand, W.-J. Chen, N. Hubert, D. Muths, G. Mou-Tham, M. Kulbicki

Tables S1-S8 and Figs. S1-S14 here appended
### Table S1 References for lists of reef fish species by locality/country in the Indian Ocean. Localities arranged by alphabetical order. General references for species lists not restricted to a particular geographic region

<table>
<thead>
<tr>
<th>Region, Region</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. Africa</td>
<td>MASDEA data base 2011</td>
</tr>
<tr>
<td></td>
<td>Harding, S., Randriamanantsoa, B., Hardy, T., Curd, A. 2006 Coral Reef Monitoring and Biodiversity Assessment to support the planning of a proposed MPA at Andavadoaka. 100 pp.</td>
</tr>
</tbody>
</table>


Samoilys and Randriamananantsoa Reef fishes of northeast Madagascar in Obura, D.O. and Oliver, T. (2011) A Rapid Marine Biodiversity Assessment of the Coral Reefs of Northeast Madagascar...

Mozambique


Nicobar


NW Australia


Oman


Red Sea


Seychelles


Socotra


South Africa


Tanzania


Thailand


W. Australia


General


Smithsonian Contribution to Zoology, 238, 1-129.


Royal Ontario Museum of Life Sciences Contributions, 159. 54+107 pp.


Table S2  Total number of species ($N$) and number of species with a limited geographic range for each biogeographic region. *Local, provincial* and *regional* ranges are defined in the main text. *All* sum of the numbers of species with limited range (also expressed as % of $N$)

<table>
<thead>
<tr>
<th>Province a</th>
<th>Abbreviation</th>
<th>$N$</th>
<th>No. species with limited geographic range</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Local</td>
<td>Provincial</td>
<td>Regional</td>
</tr>
<tr>
<td>Andaman - Cocos - Christmas</td>
<td>AND</td>
<td>1068</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Eastern African coast</td>
<td>EAF</td>
<td>1532</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>East India - Thailand - Malaysia</td>
<td>EI</td>
<td>1728</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>Maldives - Laccadives - Chagos</td>
<td>MLC</td>
<td>1410</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Northern Australia</td>
<td>NAU</td>
<td>761</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Northwestern Indian Ocean</td>
<td>NWI</td>
<td>1304</td>
<td>16</td>
<td>58</td>
</tr>
<tr>
<td>Spencer Gulf Province</td>
<td>SCI</td>
<td>612</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Southwestern Indian Ocean</td>
<td>SWI</td>
<td>1070</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>Western Australia</td>
<td>WAU</td>
<td>1931</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>Western Indian Ocean Islands</td>
<td>WII</td>
<td>1572</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>Western Indian Ocean b</td>
<td>WIO</td>
<td>1826</td>
<td>27</td>
<td>23</td>
</tr>
</tbody>
</table>

a delimited according to Kulbicki et al. (2013)
b defined by merging EAF and WII provinces
Table S3  Population-pairwise $\phi_{ST}$ estimates (Excoffier et al., 1992) for *Dascyllus abudafur / D. aruanus*. Abbreviations for samples: as in Borsa et al. (2014). Sample names arranged from West to East. WIO samples are EU, JN, MD and GL; the other Indian-Ocean sample is RS; Pacific-Ocean samples are PI, DS, TW, LL, SK, RA, MA, NC and SI. Table cells are toned grey, whose intensity is positively correlated with value of $\phi_{ST}$ estimate (ARLEQUIN 3.5; Excoffier and Lischer, 2010). Probability of $\phi_{ST}$ value under null hypothesis of genetic homogeneity estimated from 1,000 random permutations under ARLEQUIN. * P<0.05; ** P<0.01; *** P<0.001

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample</th>
<th>RS</th>
<th>EU</th>
<th>JN</th>
<th>MD</th>
<th>GL</th>
<th>PI</th>
<th>DS</th>
<th>TW</th>
<th>LL</th>
<th>SK</th>
<th>RA</th>
<th>MA</th>
<th>NC</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU</td>
<td>RS 0.354***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JN</td>
<td>0.369***</td>
<td></td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.353***</td>
<td>-0.028</td>
<td></td>
<td>0.080</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td>0.386***</td>
<td>0.000</td>
<td>-0.015</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>0.803***</td>
<td>0.703***</td>
<td>0.744***</td>
<td>0.735***</td>
<td>0.650***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS</td>
<td>0.800***</td>
<td>0.713***</td>
<td>0.744***</td>
<td>0.739***</td>
<td>0.668***</td>
<td>0.102*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>0.811***</td>
<td>0.735***</td>
<td>0.771***</td>
<td>0.759***</td>
<td>0.708***</td>
<td>0.061</td>
<td>-0.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>0.779***</td>
<td>0.703***</td>
<td>0.729***</td>
<td>0.723***</td>
<td>0.668***</td>
<td>0.026</td>
<td>-0.006</td>
<td>-0.016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>0.767***</td>
<td>0.685***</td>
<td>0.708***</td>
<td>0.705***</td>
<td>0.637***</td>
<td>0.032</td>
<td>0.021</td>
<td>0.015</td>
<td>-0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td>0.770***</td>
<td>0.693***</td>
<td>0.719***</td>
<td>0.711***</td>
<td>0.651***</td>
<td>0.079</td>
<td>0.180**</td>
<td>0.142**</td>
<td>0.090**</td>
<td>0.029</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>0.785***</td>
<td>0.676***</td>
<td>0.680***</td>
<td>0.703***</td>
<td>0.560**</td>
<td>0.202*</td>
<td>0.126</td>
<td>0.146*</td>
<td>0.104</td>
<td>0.014</td>
<td>0.042</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>0.694***</td>
<td>0.621***</td>
<td>0.625***</td>
<td>0.632***</td>
<td>0.563***</td>
<td>0.117**</td>
<td>0.060</td>
<td>0.085*</td>
<td>0.068*</td>
<td>0.020</td>
<td>0.079*</td>
<td>-0.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>0.859***</td>
<td>0.802***</td>
<td>0.836***</td>
<td>0.824***</td>
<td>0.783***</td>
<td>0.556***</td>
<td>0.549***</td>
<td>0.552***</td>
<td>0.500***</td>
<td>0.513***</td>
<td>0.547***</td>
<td>0.625***</td>
<td>0.466***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S4  Population-pairwise $\phi_{ST}$ estimates (Excoffier et al., 1992) for *Epinephelus merra*, computed using ARLEQUIN 3.5 (Excoffier and Lischer, 2010). 

**Abbreviations for samples:** NCA, New Caledonia; MRU, Moruroa atoll, Gambier archipelago; else same locations and same abbreviations as in Muths et al. (2014). Sample names arranged from West to East. WIO samples are EUR, GEY, GLO, JDN, KEN, MAD, MAY, MOH, MOR, REU, ROD, SEY and TAN; the other Indian-Ocean sample is MAL; Pacific-Ocean samples are NCA and MRU. Table cells are toned grey, with intensity of grey tone positively correlated with value of $\phi_{ST}$ estimate. Probability of $\phi_{ST}$ value under null hypothesis of genetic homogeneity estimated from 1,000 random permutations under ARLEQUIN. * $P<0.05$; ** $P<0.01$; *** $P<0.001$

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample</th>
<th>TAN</th>
<th>KEN</th>
<th>EUR</th>
<th>JDN</th>
<th>MOR</th>
<th>MOH</th>
<th>MAY</th>
<th>GEY</th>
<th>GLO</th>
<th>MAD</th>
<th>SEY</th>
<th>REU</th>
<th>MAU</th>
<th>ROD</th>
<th>MAL</th>
<th>NCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEN</td>
<td></td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR</td>
<td>-0.032</td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JDN</td>
<td>-0.030</td>
<td>0.065</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOR</td>
<td>0.018</td>
<td>0.214**</td>
<td>0.046*</td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOH</td>
<td>-0.033</td>
<td>0.031</td>
<td>-0.013</td>
<td>-0.015</td>
<td>0.039*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAY</td>
<td>-0.026</td>
<td>0.170</td>
<td>0.010</td>
<td>-0.019</td>
<td>-0.060</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEY</td>
<td>-0.032</td>
<td>0.080</td>
<td>0.002</td>
<td>-0.013</td>
<td>0.015</td>
<td>-0.007</td>
<td>-0.033</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLO</td>
<td>-0.033</td>
<td>0.084*</td>
<td>0.000</td>
<td>-0.012</td>
<td>0.007</td>
<td>-0.008</td>
<td>-0.038</td>
<td>-0.016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>0.077*</td>
<td>0.307***</td>
<td>0.090**</td>
<td>0.055*</td>
<td>0.030</td>
<td>0.083***</td>
<td>-0.057</td>
<td>0.034*</td>
<td>0.036*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEY</td>
<td>-0.050</td>
<td>0.057</td>
<td>-0.006</td>
<td>-0.017</td>
<td>0.022</td>
<td>-0.011</td>
<td>-0.026</td>
<td>-0.020</td>
<td>-0.017</td>
<td>0.056*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REU</td>
<td>0.015</td>
<td>0.198**</td>
<td>0.059*</td>
<td>0.033</td>
<td>0.005</td>
<td>0.057*</td>
<td>-0.053</td>
<td>0.011</td>
<td>0.018</td>
<td>-0.011</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAU</td>
<td>-0.026</td>
<td>0.109*</td>
<td>0.009</td>
<td>-0.009</td>
<td>0.000</td>
<td>0.004</td>
<td>-0.051</td>
<td>-0.012</td>
<td>-0.011</td>
<td>0.015</td>
<td>-0.018</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROD</td>
<td>0.102*</td>
<td>0.344***</td>
<td>0.128***</td>
<td>0.086**</td>
<td>0.022</td>
<td>0.120***</td>
<td>-0.031</td>
<td>0.068**</td>
<td>0.063***</td>
<td>0.002</td>
<td>0.075*</td>
<td>-0.006</td>
<td>0.042*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAL</td>
<td>0.189**</td>
<td>0.006</td>
<td>0.161***</td>
<td>0.222***</td>
<td>0.369***</td>
<td>0.178***</td>
<td>0.356***</td>
<td>0.240***</td>
<td>0.242***</td>
<td>0.445***</td>
<td>0.230***</td>
<td>0.368***</td>
<td>0.274***</td>
<td>0.480***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCA</td>
<td>-0.016</td>
<td>0.069</td>
<td>0.024</td>
<td>0.026</td>
<td>0.050</td>
<td>0.025</td>
<td>-0.005</td>
<td>0.016</td>
<td>0.018</td>
<td>0.093*</td>
<td>0.025</td>
<td>0.054</td>
<td>0.023</td>
<td>0.138**</td>
<td>0.247**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRU</td>
<td>-0.072</td>
<td>-0.140</td>
<td>-0.029</td>
<td>0.023</td>
<td>0.199</td>
<td>0.003</td>
<td>0.224</td>
<td>0.043</td>
<td>0.042</td>
<td>0.336*</td>
<td>0.018</td>
<td>0.168</td>
<td>0.060</td>
<td>0.371</td>
<td>0.076</td>
<td>-0.201</td>
<td></td>
</tr>
</tbody>
</table>
Table S5  Population-pairwise $\phi_{ST}$ estimates (Excoffier et al., 1992) for *Lutjanus kasmira*, computed using ARLEQUIN 3.5 (Excoffier and Lischer, 2010). Samples Christmas Island (*CHR*), Cocos-Keeling (*COC*), Diego Garcia (*DIE*), Fiji (*FIJ*), Guam (*GUA*), Kirimitati (*KIR*), Lizard Island (*LIZ*), Marquesas (*MAR*), Moorea (*MOO*), Okinawa (*OKI*), Seychelles (*SEC*) and Sodwana Bay (*SOD*) from Gaither et al. (2010); other abbreviations: NOS, Nosy Be (Muths et al., 2012), else as in Muths et al. (2014). Sample names arranged from West to East. WIO samples are SOD, TAN, KEN, EUR, JDN, MOR, MOH, MAY, GEY, GLO, NOS, MAD, SEC, SEY, REU, MAU and ROD; the other Indian-Ocean samples are DIE, MAL, COC and CHR; Pacific-Ocean samples are OKI, LIZ, GUA, FIJ, KIR, MOO and MAR. Table cells are toned grey, with intensity of grey tone positively correlated with value of $\phi_{ST}$ estimate. Probability of $\phi_{ST}$ value under null hypothesis of genetic homogeneity estimated from 1,000 random permutations under ARLEQUIN. *P<0.05; **P<0.01; ***P<0.001

<table>
<thead>
<tr>
<th>Sample</th>
<th>SOD</th>
<th>TAN</th>
<th>KEN</th>
<th>EUR</th>
<th>JDN</th>
<th>MOR</th>
<th>MOH</th>
<th>MAY</th>
<th>GEY</th>
<th>GLO</th>
<th>NOS</th>
<th>MAD</th>
<th>SEC</th>
<th>SEY</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAN</td>
<td>0.016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEN</td>
<td>0.016</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR</td>
<td>0.006</td>
<td>0.021</td>
<td>-0.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JDN</td>
<td>0.000</td>
<td>0.006</td>
<td>-0.007</td>
<td>-0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOR</td>
<td>-0.005</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.002</td>
<td>-0.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOH</td>
<td>0.001</td>
<td>0.009</td>
<td>0.009</td>
<td>0.001</td>
<td>-0.010</td>
<td>-0.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAY</td>
<td>-0.009</td>
<td>0.007</td>
<td>-0.012</td>
<td>-0.008</td>
<td>-0.005</td>
<td>-0.005</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEY</td>
<td>0.008</td>
<td>0.014</td>
<td>0.005</td>
<td>0.005</td>
<td>-0.010</td>
<td>-0.008</td>
<td>-0.005</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLO</td>
<td>-0.005</td>
<td>0.002</td>
<td>-0.007</td>
<td>-0.006</td>
<td>-0.013</td>
<td>-0.014</td>
<td>-0.011</td>
<td>-0.006</td>
<td>-0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOS</td>
<td>-0.006</td>
<td>-0.010</td>
<td>0.002</td>
<td>0.011</td>
<td>-0.006</td>
<td>-0.013</td>
<td>-0.011</td>
<td>0.003</td>
<td>0.000</td>
<td>-0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>-0.014</td>
<td>-0.003</td>
<td>0.008</td>
<td>-0.004</td>
<td>-0.014</td>
<td>-0.023</td>
<td>-0.027</td>
<td>-0.007</td>
<td>-0.003</td>
<td>-0.019</td>
<td>-0.028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC</td>
<td>-0.007</td>
<td>0.000</td>
<td>-0.026</td>
<td>-0.004</td>
<td>-0.021</td>
<td>-0.020</td>
<td>-0.015</td>
<td>-0.010</td>
<td>-0.021</td>
<td>-0.020</td>
<td>-0.014</td>
<td>0.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEY</td>
<td>-0.011</td>
<td>-0.028</td>
<td>0.007</td>
<td>-0.019</td>
<td>-0.019</td>
<td>-0.010</td>
<td>-0.006</td>
<td>0.003</td>
<td>-0.014</td>
<td>-0.028</td>
<td>-0.026</td>
<td>-0.026</td>
<td>-0.031</td>
<td></td>
</tr>
<tr>
<td>RUN</td>
<td>0.004</td>
<td>0.010</td>
<td>0.008</td>
<td>0.008</td>
<td>-0.012</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.003</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.002</td>
<td>0.008</td>
<td>-0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>MAU</td>
<td>0.042*</td>
<td>0.061*</td>
<td>0.051*</td>
<td>0.053**</td>
<td>0.055***</td>
<td>0.047**</td>
<td>0.058***</td>
<td>0.037**</td>
<td>0.051**</td>
<td>0.055***</td>
<td>0.046**</td>
<td>0.043</td>
<td>0.054*</td>
<td>0.061*</td>
</tr>
<tr>
<td>ROD</td>
<td>-0.005</td>
<td>0.019</td>
<td>0.002</td>
<td>0.003</td>
<td>0.002</td>
<td>0.004</td>
<td>0.006</td>
<td>-0.002</td>
<td>0.015</td>
<td>0.008</td>
<td>0.004</td>
<td>0.005</td>
<td>0.004</td>
<td>0.009</td>
</tr>
<tr>
<td>DIE</td>
<td>-0.006</td>
<td>0.002</td>
<td>0.014</td>
<td>-0.001</td>
<td>-0.006</td>
<td>-0.004</td>
<td>0.001</td>
<td>0.002</td>
<td>-0.003</td>
<td>-0.023</td>
<td>-0.024</td>
<td>-0.016</td>
<td>-0.026</td>
<td></td>
</tr>
<tr>
<td>MAL</td>
<td>-0.014</td>
<td>-0.016</td>
<td>-0.035</td>
<td>-0.015</td>
<td>-0.012</td>
<td>-0.018</td>
<td>-0.004</td>
<td>-0.022</td>
<td>-0.008</td>
<td>-0.018</td>
<td>-0.006</td>
<td>-0.011</td>
<td>-0.022</td>
<td>-0.028</td>
</tr>
<tr>
<td>COC</td>
<td>0.003</td>
<td>0.002</td>
<td>0.006</td>
<td>0.000</td>
<td>0.003</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.003</td>
<td>0.002</td>
<td>0.009</td>
<td>0.002</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>CHR</td>
<td>0.011</td>
<td>0.023</td>
<td>0.006</td>
<td>0.024</td>
<td>0.007</td>
<td>0.006</td>
<td>0.019</td>
<td>0.005</td>
<td>0.011</td>
<td>0.010</td>
<td>0.010</td>
<td>0.023</td>
<td>-0.006</td>
<td>-0.002</td>
</tr>
<tr>
<td>OKI</td>
<td>-0.030</td>
<td>-0.292</td>
<td>0.331</td>
<td>-0.317</td>
<td>-0.317</td>
<td>-0.308</td>
<td>-0.313</td>
<td>-0.305</td>
<td>-0.311</td>
<td>-0.317</td>
<td>-0.298</td>
<td>-0.305</td>
<td>-0.327</td>
<td>-0.315</td>
</tr>
<tr>
<td>LIZ</td>
<td>-0.066</td>
<td>-0.058</td>
<td>-0.069</td>
<td>-0.054</td>
<td>-0.070</td>
<td>-0.073</td>
<td>-0.057</td>
<td>-0.076</td>
<td>-0.051</td>
<td>-0.069</td>
<td>-0.064</td>
<td>-0.026</td>
<td>-0.071</td>
<td>-0.082</td>
</tr>
<tr>
<td>GUU</td>
<td>0.016*</td>
<td>0.303*</td>
<td>0.154*</td>
<td>0.061</td>
<td>0.011</td>
<td>0.015*</td>
<td>0.007</td>
<td>0.012</td>
<td>0.009</td>
<td>0.023**</td>
<td>0.021</td>
<td>0.030</td>
<td>0.007</td>
<td>0.024</td>
</tr>
<tr>
<td>FIJ</td>
<td>0.023*</td>
<td>0.026</td>
<td>0.191</td>
<td>0.015*</td>
<td>0.004</td>
<td>0.007</td>
<td>0.012</td>
<td>0.010</td>
<td>0.009</td>
<td>0.005</td>
<td>0.021*</td>
<td>0.030</td>
<td>0.007</td>
<td>0.024</td>
</tr>
<tr>
<td>KIR</td>
<td>0.032**</td>
<td>0.039</td>
<td>0.024</td>
<td>0.032***</td>
<td>0.015*</td>
<td>0.021**</td>
<td>0.027**</td>
<td>0.018*</td>
<td>0.023*</td>
<td>0.021***</td>
<td>0.034***</td>
<td>0.030</td>
<td>0.018</td>
<td>0.028*</td>
</tr>
<tr>
<td>MOO</td>
<td>0.040***</td>
<td>0.066**</td>
<td>0.046*</td>
<td>0.031***</td>
<td>0.036***</td>
<td>0.033</td>
<td>0.036***</td>
<td>0.028</td>
<td>0.036***</td>
<td>0.035***</td>
<td>0.047</td>
<td>0.026*</td>
<td>0.042**</td>
<td>0.063</td>
</tr>
<tr>
<td>MAR</td>
<td>0.536*</td>
<td>0.569***</td>
<td>0.558***</td>
<td>0.590*</td>
<td>0.597***</td>
<td>0.560***</td>
<td>0.548***</td>
<td>0.554***</td>
<td>0.575***</td>
<td>0.551***</td>
<td>0.578***</td>
<td>0.639***</td>
<td>0.635***</td>
<td></td>
</tr>
</tbody>
</table>

Table continued...
Table S6 Population-pairwise $\phi_{ST}$ estimates (Excoffier et al., 1992) for *Panulirus homarus*, computed using ARLEQUIN 3.5 (Excoffier and Lischer, 2010). Samples were constituted as following. *WIO* included samples Za and Tn of Lavery et al. (2014) excluding 3 haplotypes of Za assigned to *P. h. rubellus*; *NWIO* included samples Om, L and Ch of Lavery et al. (2014); *Kollam* included sample QLN of Jeena (2013); *SriLanka* included samples JSPHK and JSPHW of Senevirathna and Munasinghe (2014) (GenBank nos. KC959890, KC959891 and KF548571) and sequences nos. KF715528-KF715550 and KF715552 from GenBank; *Chennai* is sample CHE of Jeena (2013); *Visakhapatnam* included sample VSK of Jeena (2013) and sequence no. JN418937 from GenBank; *SChinaSea* included samples Vn and Tw of Lavery et al. (2014) and sequences nos. JN542716 and JN591360 from GenBank; *Marquesas* is sample M of Lavery et al. (2014). Sample names arranged from West to East. Indian-Ocean samples other than WIO are: NWIO, Kollam, SriLanka, Chennai and Visakhapatnam; Pacific-Ocean samples are SchinaSea and Marquesas. Table cells are toned grey, with intensity of grey tone positively correlated with value of $\phi_{ST}$ estimate. Probability of $\phi_{ST}$ value under null hypothesis of genetic homogeneity estimated from 1,000 random permutations under ARLEQUIN. * $P<0.05$; ** $P<0.01$; *** $P<0.001$

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample</th>
<th>WIO</th>
<th>NWIO</th>
<th>Kollam</th>
<th>SriLanka</th>
<th>Chennai</th>
<th>Visakhapatnam</th>
<th>SChinaSea</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWIO</td>
<td></td>
<td>-0.284***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kollam</td>
<td></td>
<td>0.025</td>
<td>0.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SriLanka</td>
<td></td>
<td>0.014</td>
<td>0.041</td>
<td></td>
<td>-0.037</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chennai</td>
<td></td>
<td>-0.013</td>
<td>0.127**</td>
<td>-0.033</td>
<td></td>
<td>-0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visakhapatnam</td>
<td></td>
<td>0.068</td>
<td>0.064**</td>
<td>-0.011</td>
<td>0.016</td>
<td>0.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SChinaSea</td>
<td></td>
<td>-0.059</td>
<td>0.176***</td>
<td>0.024</td>
<td>0.035</td>
<td>0.034</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>Marquesas</td>
<td></td>
<td>0.621**</td>
<td>0.749***</td>
<td>0.610***</td>
<td>0.525***</td>
<td>0.619***</td>
<td>0.639***</td>
<td>0.531***</td>
</tr>
</tbody>
</table>
Table S7 Population-pairwise $\phi_{ST}$ estimates (Excoffier et al., 1992) for *Terapon jarbua*, computed using ARLEQUIN 3.5 (Excoffier and Lischer, 2010) on the basis of partial CO1 gene sequences. Samples were constituted as following. *Safirica* included four individual sequences from Kwazulu-Natal (GenBank nos. JF494663-JF494666); *RedSea* included two individual sequences from the Gulf of Eilat (GenBank nos. KJ466137, KJ466138); *Eparses* included seven new sequences from Juan de Nova; *Yemen* included all samples from Al-Mukalla and Socotra in Lavergne et al. (2014) (GenBank nos. JQ342095-JQ342111); *Gulf* included three individual sequences from Bushehr, Persian Gulf (Asgharian et al., 2011) (GenBank nos. HQ149959-HQ149961); *Windia* included a sequence published by Lakra et al. (2011) (GenBank no. FJ237549) and three individual sequences from the western coast of the Indian peninsula (GenBank nos. KC774674, KF268188 and KF268189); *Eindia* included four individual sequences from the eastern coast of the Indian peninsula (GenBank nos. FJ265859, JX260979, KC241987 and KC417308); *SchinaSea* included two sequences from off southeastern China (GenBank nos. EU871691, EU871692), three new sequences from around Nha Trang, Vietnam and 10 new sequences from around Kenting, Taiwan; *Philippines* included four individual sequences from the Philippines archipelago (GenBank nos. JN021254, JN021255, KC970423, KF009671). Sample names arranged from West to East. WIO samples are Safirica and RedSea; Indian-Ocean samples other than WIO are: RedSea, Yemen, Gulf, Windia and Eindia; Pacific-Ocean samples are SchinaSea and Philippines. Table cells are toned grey, with intensity of grey tone positively correlated with value of $\phi_{ST}$ estimate. Probability of $\phi_{ST}$ value under null hypothesis of genetic homogeneity estimated from 1,000 random permutations under ARLEQUIN. * P<0.05; ** P<0.01; *** P<0.001

<table>
<thead>
<tr>
<th>Sample</th>
<th>Safirica</th>
<th>RedSea</th>
<th>Eparses</th>
<th>Yemen</th>
<th>Gulf</th>
<th>Windia</th>
<th>Eindia</th>
<th>SchinaSea</th>
</tr>
</thead>
<tbody>
<tr>
<td>RedSea</td>
<td>-0.336</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eparses</td>
<td>-0.197</td>
<td>0.363</td>
<td>0.613***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yemen</td>
<td>0.655**</td>
<td>0.363</td>
<td></td>
<td>0.517*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gulf</td>
<td>0.571</td>
<td>0.205</td>
<td>0.500*</td>
<td>0.084</td>
<td>0.067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windia</td>
<td>0.543</td>
<td>0.149</td>
<td>0.472*</td>
<td>0.057</td>
<td>-0.044</td>
<td>-0.243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eindia</td>
<td>0.508</td>
<td>0.099</td>
<td>0.472*</td>
<td>0.057</td>
<td>-0.044</td>
<td>-0.243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SchinaSea</td>
<td>0.676***</td>
<td>0.423</td>
<td>0.628***</td>
<td>0.081*</td>
<td>0.004</td>
<td>-0.030</td>
<td>-0.045</td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>0.565*</td>
<td>0.204</td>
<td>0.506*</td>
<td>0.279**</td>
<td>0.329</td>
<td>-0.101</td>
<td>-0.088</td>
<td>0.056</td>
</tr>
<tr>
<td>Species</td>
<td>Pelagic larval duration (days)</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------------</td>
<td>------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cephalopholis argus</em></td>
<td>25-29 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chaetodon ornatissimus</em></td>
<td>61-64 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Dascyllus aruanus</em></td>
<td>18.6-24.3 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Epinephelus areolatus</em></td>
<td>18 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Epinephelus merra</em></td>
<td>29.6 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fistularia commersonii</em></td>
<td>42-48 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Labroides dimidiatus</em></td>
<td>23.2-24.7 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lethrinus nebulosus</em></td>
<td>≥37 d</td>
<td>Brothers et al. (1983)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Linckia laevigata</em></td>
<td>≥22 d</td>
<td>Yamaguchi (1973) in Crandall et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lutjanus kasmira</em></td>
<td>31.0 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Naso lituratus</em></td>
<td>65-72 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Panulirus argus</em></td>
<td>174 ± 22 d</td>
<td>Goldstein et al. (2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Panulirus homarus</em></td>
<td>ca. 180 d</td>
<td>Phillips and Matsuda (2011) in Lavery et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Scarus schlegeli</em></td>
<td>28-35 d</td>
<td>Stier et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Tridacna maxima</em></td>
<td>10-19 d</td>
<td>Jameson (1976)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. S1 Aggregation of localities in the Indian Ocean, defined by their checklists of reef fishes (from references in Table S1). Inset: hierarchical classification based on Ward’s aggregation algorithm (Legendre and Legendre, 2012) using Pearson’s r coefficient (A), Euclidean distance (B), and squared Euclidean distance (C). Different clusters are distinguished by different colours.
**Fig. S2**  Size frequency distribution of reef fish species according by biogeographic province (see Table S2 for abbreviations and Fig. 1 of main article for province boundaries). *Abcissa*, body size (in cm); *ordinates*, frequency (%).
Fig. S3  Phylogeographic structure in peacock grouper, *Cephalopholis argus* (Gaither et al., 2011). Samples encoded according to geographic origin: black, western Indian Ocean (WIO); grey, Indian Ocean; white, Pacific Ocean. Gaither et al.’s (2011) sample from Bali was arbitrarily placed in the Indian Ocean. A  Parsimony network of cytochrome *b* gene haplotypes (modified from Gaither et al., 2011). Two major haplogroups were observed, with a clear Indian vs. Pacific phylogeographic partition. This indicates separate evolution between an ‘Indian’ and a ‘Pacific’ mitochondrial forms. However, the ‘Indian’ haplogroup also harboured haplotypes sampled in the Pacific Ocean and vice versa, suggesting either incomplete lineage sorting, or secondary overlap between the two oceans. The latter hypothesis is the one favoured by Gaither et al. (2011). Haplotypes sampled in the WIO were nested within the ‘Indian’ haplogroup exclusively, indicating that ‘Pacific’ haplotypes under Gaither et al.’s (2011) overlap hypothesis have not reached the WIO. B  Sampling locations.
**Fig. S4** Scrawled butterflyfish, *Chaetodon meyeri*. **A** Parsimony network (redrawn from DiBattista et al., 2012) of the cytochrome *b* gene sequences of 134 individuals from various localities in the Indo-West Pacific, including the WIO (black), the rest of the Indian Ocean (grey), and the Western Pacific (white). Scale bar: one mutational step. Only moderate Indian vs. Pacific phylogeographic partition was observed, suggesting weak differentiation between populations from the two oceans. **B** Sampling locations.
Fig. S5 Phylogeographic structure in cleaner wrasse, *Labroides* sp. 1 (Sims et al., 2014). Samples encoded according to geographic origin: black, western Indian Ocean; grey, Indian Ocean; white, Pacific Ocean. 

A Median-joining parsimony network (Bandelt et al., 1999) of mitochondrial control-region haplotypes. Sixty three sequences were selected (GenBank accession nos. EU256890-EU256903, KC151127-KC15138, KC151172-KC151184, KC151212, KC151214-KC151216, KC151219, KC151227, KC151230, KC151232, KC151234, KC151236-KC151250), aligned under BioEdit (Hall, 1999) and trimmed to 267 base pairs. Scale bar, one mutational step. Loose phylogeographic structure was inferred from the structure of the network, where Pacific haplotypes tended to cluster at one side of the network and Indian haplotypes, including those from the WIO, tended to cluster on the other side. A few haplotypes from the ‘Indian’ cluster were sampled in the Pacific Ocean and vice versa, suggesting either incomplete lineage sorting, or secondary gene flow between populations from the two oceans. This pattern is also compatible with isolation by distance (Wright, 1942), the hypothesis favoured by Sims et al. (2014) on the basis of observed correlation between $\phi_{st}$ and geographic distance. 

B Sampling locations (redrawn from Sims et al., 2014).
Fig. S6 Longface emperor, *Lethrinus* sp. A sensu Borsa et al. (2013b). A Median Joining parsimony network (Bandelt et al., 1999) of the cytochrome *b* gene sequences of 16 individuals from various localities in the Indo-West Pacific, including the WIO (black), the rest of the Indian Ocean (grey), and the Western Pacific (white). Haplotype sequences are from Fig. 5 of Borsa et al. (2013b). Complete sequences have nos. AF381252, KC596033-KC596044, KC596046, KC596048, and KC596051 in GENBANK. Scale bar: one mutational step; number indicates the number of mutations separating the two main haplogroups. A clear Indian vs. Pacific phylogeographic partition was observed, indicating a geographic barrier between the two mitochondrial forms. The geographic barrier is located east of Bali and west of West Papua. B Sampling locations.
Fig. S7 Blue sea star, *Linckia laevigata*. A Parsimony network (redrawn from Crandall et al., 2014) of the partial *CO1* gene sequences of 791 individuals from various localities in the Indo-West Pacific, including the WIO (black), the eastern Indian Ocean (grey), and the western Pacific Ocean (white). Scale bar: one mutational step. Two major clusters of haplotypes were observed, both including eastern Indian and Pacific haplotypes. One of the two clusters, which included a few WIO haplotypes, also had a substantially higher proportion of Indian-Ocean haplotypes than the other cluster, which was predominantly Pacific. This suggests either incomplete lineage sorting between populations from the Indian vs. Pacific oceans, or secondary admixture of haplotypes from one ocean to the other and vice versa. Interestingly, the dominant WIO haplotypes occupied a central position in the network, implying that one or the other, possibly both, of the two major haplogroups may derive from an ancestral WIO stock, or from a geographically larger region encompassing the WIO. B Sampling locations.
Fig. S8 Phylogeographic structure in spotted unicornfish, *Naso brevirostris*. Samples encoded according to geographic origin: *black*, western Indian Ocean (WIO); *grey*, Indian Ocean; *white*, Pacific Ocean.  A Median-joining parsimony network (Bandelt et al., 1999) of mitochondrial control-region haplotypes. One hundred two sequences were retrieved from GenBank (nos. FJ216727-FJ216828) and aligned under BioEdit (Hall, 1999). Gaps of one to five nucleotides were inserted; sequences had a common length of 247 base pairs. Haplotypes were organized into two loose haplogroups. Each haplogroup possessed haplotypes from the WIO, the rest of the Indian Ocean, and the Pacific Ocean and no geographic differences were observed in haplogroup frequencies (Horne et al., 2008).  *Scale bar*, five mutational steps.  B Sampling locations.
Fig. 59 Phylogeographic structure in bluespine unicornfish, _Naso unicornis_. Samples encoded according to geographic origin: _black_, western Indian Ocean; _grey_, Indian Ocean; _white_, Pacific Ocean. A Median-joining parsimony network (Bandelt et al., 1999) of mitochondrial control-region haplotypes. One hundred seven sequences (Horne et al., 2008) were retrieved from [GenBank](https://www.ncbi.nlm.nih.gov/genbank/) (nos. FJ216829-FJ216935) and aligned under [BioEdit](https://www.mbio.ncsu.edu/BioEdit/bioedit.html) (Hall, 1999). Gaps of one to two nucleotides were inserted; sequences had a common length of 277 base pairs. No phylogeographic structure was detected, either from this network or from $\phi_{ST}$ analysis (Horne et al., 2008). Scale bar, one mutational step. B Sampling locations.
Fig. S10 Phylogeographic structure in bignose unicornfish, *Naso vlamingii*. Samples encoded according to geographic origin: *black*, western Indian Ocean; *grey*, Indian Ocean; *white*, Pacific Ocean. **A** Median-joining parsimony network (Bandelt et al., 1999) of mitochondrial control-region haplotypes. One hundred thirteen sequences were retrieved from GENBANK (nos. DQ767974-DQ768086) and aligned under BIOEDIT (Hall, 1999). Gaps of one to five nucleotides were inserted; sequences had a common length of 354 base pairs. Haplotypes were organized into four loose haplogroups, coined “Clades I-IV” by Klanten et al. (2007) (I-IV here). Each haplogroup possessed haplotypes from the WIO, the rest of the Indian Ocean, and the Pacific Ocean and no geographic differences were observed in haplogroup frequencies (Klanten et al., 2007). **Scale bar**, five mutational steps. **B** Sampling locations (redrawn from Klanten et al., 2007).
Fig. S8 Blue-spotted maskray, *Neotrygon kuhlii*. A Parsimony network (edited from Arlyza et al., 2013) of the partial CO1 gene sequences of 147 individuals from various localities in the Indo-West Pacific, including the WIO (black), the eastern Indian Ocean (grey), and the western Pacific Ocean (white). Scale bar: one mutational step. Eight distinct haplogroups were observed, including haplogroups II-VIII from the Coral Triangle region, and a loosely defined haplogroup (Ia, Ib) from the Indian Ocean (Arlyza et al., 2013). Coral-Sea haplotypes not represented here, as they pertain to the recently-resurrected *N. trigonoides* (see Borsa et al., 2013a). B Sampling locations.
Fig. S12 Phylogeographic structure in common parrotfish, *Scarus psittacus*. Samples encoded according to geographic origin: *black*, western Indian Ocean; *grey*, Indian Ocean; *white*, Pacific Ocean. **A** Median-joining parsimony network (Bandelt et al., 1999) of mitochondrial control-region haplotypes. One hundred sixty nine sequences were retrieved from GenBank (nos. AY324601, EU926978-EU927144 and JX026573), aligned under BioEdit (Hall, 1999) and trimmed to 270 base pairs. *Scale bar*, one mutational step. Haplotypes were organized as a quasi-star like network with a dominant central haplotype that was shared between the WIO, the central and eastern Indian Ocean, and the Pacific Ocean. $\phi_{ST}$ analysis revealed genetic differences between the WIO population from Seychelles and the eastern-Indian Ocean populations, which was thought to reflect peripheral genetic differentiation (Winters et al., 2010). **B** Sampling locations (modified from Winters et al., 2010).
Fig. S13 Phylogeographic structure in whitetip reef shark, *Triaenodon obesus* (Whitney et al., 2012). Samples encoded according to geographic origin: **black**, western Indian Ocean; **grey**, Indian Ocean; **white**, Pacific Ocean. 

**A** Median-joining parsimony network (Bandelt et al., 1999) of mitochondrial control-region haplotypes, aligned under BioEdit (Hall, 1999) and reduced to a common length of 881 bp. *Scale bar*, one mutational step. No obvious Indian vs. Pacific partition was visible from the network. However, this apparent lack of phylogeographic structure may be an artefact of unbalanced sampling design. $\phi_{ST}$ analysis revealed substantial genetic differentiation between populations either side of the Indo-Pacific barrier (Whitney et al., 2012). The only WIO haplotype, from the Seychelles, was identical to the haplotype dominant in the Indian Ocean. 

**B** Sampling locations (redrawn from Whitney et al., 2014). Two additional Pacific samples, not represented on this figure, were from Costa Rica.
Phylogeographic structure in small giant clam, *Tridacna maxima*. Samples encoded according to geographic origin: black, western Indian Ocean; grey, Indian Ocean; white, Pacific Ocean. A Parsimony network of COI haplotypes (GenBank accession nos. FM244476-FM244619 and HE995454-HE995487; N = 151) (redrawn from Hui, 2012). Seven major haplogroups, coined clades 1-7 by Hui (2012) were observed, each geographically circumscribed. Clade 1 was the single one represented in samples from eastern Indonesia (dotted ellipse); it was also marginally present in samples Karimunjawa (Java Sea, Indonesia) and in Biak (northern West Papua); clade 2 was exclusively present in the Andaman Sea and in western Indonesia; clade 3 was exclusively present in Biak; clade 4 was exclusive to the Red Sea; clades 5 and 6 were exclusively sampled from Kenya, WIO; clade 7 was exclusively sampled in French Polynesia. Strong phylogeographic partition was thus revealed. Maximum-likelihood phylogenetic analysis (T92+G+I; MEGA6; Tamura et al., 2013) using homologous sequences from *T. crocea*, *T. derasa*, *T. gigas*, *T. noae* and *T. squamosa* (GenBank accession nos. EU341379, GQ166591, KJ202113, KC456023 and EU346364, respectively) as outgroups determined clades 3 and 7 to be outgroups and placed the root at the basis of a rattle comprising clade 5, clades (4+6), and the rest. Scale bar, one mutational step. B Sampling locations (from Hui, 2012). Bk sample Biak; Ka sample Karimunjawa; dotted ellipse delineates all samples from eastern Indonesia.
References (other than Table S1)


