Received Date : 13-Dec-2014

Revised Date : 14-Nov-2015

Accepted Date : 09-Dec-2015

Article type : Special Issue

Original Article

LRH: P. A. Ahti *et al*.

RRH: Phylogeography of Indo-Pacific coral reef fishes

Phylogeography of Indo-Pacific reef fishes: sister wrasses *Coris gaimard* and *C. cuvieri* in the Red Sea, Indian Ocean and Pacific Ocean

Pauliina A. Ahti[1,2*], Richard R. Coleman[1], Joseph D. DiBattista[3,4], Michael L. Berumen[3], Luiz A. Rocha[5] and Brian W. Bowen[1]

[1]*Hawai'i Institute of Marine Biology, University of Hawai'i, Kāne'ohe, HI 96744, USA,* [2]*School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8QQ, Scotland,* [3]*Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia,* [4]*Department of Environment and Agriculture, Curtin University, PO Box U1987, Perth, WA 6845, Australia ,* [5]*Section of Ichthyology, California Academy of Sciences, San Francisco, CA 94118, USA*

*Correspondence: Pauliina A. Ahti, School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8QQ, Scotland, E-mail: pauliina.ahti@gmail.com

Word count: 7630

ABSTRACT

Aim We aim to resolve evolutionary history, biogeographic barriers and population histories for sister species of wrasses, the African Coris (*Coris cuvieri*) in the Indian Ocean and Red Sea, and the Yellowtail Coris (*Coris gaimard*) in the Pacific Ocean. Glacial sea level fluctuations during the Pleistocene have shaped the evolutionary trajectories of Indo-Pacific marine fauna, primarily by creating barriers between the Red Sea, Indian Ocean and Pacific Ocean. Here we evaluate the influence of these episodic glacial barriers on sister species *C. cuvieri* and *C. gaimard*.

Location Red Sea, Indian Ocean, Pacific Ocean.

Methods Sequences from mitochondrial DNA cytochrome oxidase *c* subunit I (COI), and nuclear introns gonadotropin-releasing hormone (GnRH) and ribosomal S7 protein were analysed in 426 individuals from across the range of both species. Median-joining networks, analysis of molecular variance (AMOVA) and Bayesian estimates of the time since most recent common ancestor were used to resolve recent population history and connectivity.

Results COI haplotypes showed a divergence of 0.97% between species, and nuclear alleles were shared between species. No population structure was detected between the Indian Ocean and Red Sea. The strongest signal of population structure was in *C. gaimard* between the Hawaiian biogeographic province and other Pacific locations (COI ϕ_{ST} = 0.040 to 0.173, $P < 0.006$; S7 ϕ_{ST} = 0.046, $P < 0.001$; GnRH ϕ_{ST} = 0.022, $P < 0.005$). Time to most recent common ancestor is approximately 2.12 Ma for *C. cuvieri* and 1.76 Ma for *C. gaimard.*

Main Conclusions We demonstrate an Indian-Pacific divergence of approximately 2 Myr and high contemporary gene flow between the Red Sea and Indian Ocean, mediated in part by the long pelagic larval stage. The discovery of hybrids at Christmas Island indicates that Indian and Pacific lineages have come into secondary contact after allopatric isolation. Subspecies status may be appropriate for these two wrasses.

Key Words Christmas Island, Hawaiian Archipelago, hybridisation, introns, Labridae, marine biogeography, mimicry, mtDNA, Red Sea

INTRODUCTION

Reef fishes are relatively sedentary as juveniles and adults, and dispersal usually occurs during the larval stage (Sale, 1980). Pelagic larval duration (PLD) may influence dispersal and population structure (Faurby & Barber, 2012; Selkoe *et al*., 2014). However, even closely related species with similar life histories may have markedly different population structure (Rocha *et al*., 2002; DiBattista *et al*., 2012). Phylogeographical surveys across the Indo-Pacific have shown that some reef fishes exhibit high genetic connectivity across oceans (Reece *et al*., 2011), while others show strong genetic structuring within relatively small areas (Timm & Kochzius, 2008).

The nature of these genetic partitions is subject to much debate, although many align with biogeographical barriers defined by endemism (Briggs & Bowen, 2013). One such barrier is the shallow and narrow Strait of Bab al Mandab between the Red Sea and Indian Ocean. Sea level fluctuations during glacial cycles decreased the water flow through this strait, contributing to the high levels of endemism in the Red Sea (Siddal *et al*., 2003; DiBattista *et al*., 2015). Cold water upwelling outside the Red Sea, from Somalia to the Arabian Peninsula, likely increases this isolation (Kemp, 1998). At the same time, the East African Coastal Current and the Somali Current along the east coast of Africa cause seasonal mixing of waters from the Red Sea and Western Indian Ocean (Obura, 2012). Another barrier lies between the Indian and Pacific Oceans, where low sea levels exposed the Sunda and Sahul shelves that connect South East Asia and Australia (Ludt & Rocha, 2015). This intermittent Indo-Pacific Barrier (IPB) has restricted gene flow between the Indian and the Pacific Oceans and has had a strong influence on the evolutionary history of Indo-Pacific species (Randall, 1998; Barber *et al*., 2006; Liu *et al*., 2014).

In addition to these two geographical barriers, oceanic distances may impose substantial barriers for shallow reef fauna. Hawai'i is the most isolated island group in the world, with the highest level of endemism for fishes at 25% (Randall, 2007). To the south of Hawai'i, the open ocean barrier is broken by smaller island groups, such as the Line Islands and Johnston Atoll. Although remote, these islands may act as 'stepping stones' for colonization of Hawai'i (Hourigan & Reese, 1987). The Kuroshio Current flowing north-east from southern Japan is thought to be another dispersal corridor for larvae to reach the remote Hawaiian Archipelago (Randall, 1998; Bird *et al.,* 2011).

The relationship between marine biogeographical barriers and evolutionary patterns can be illuminated with sister species, the products of recent evolutionary bifurcations. Examinations of the youngest phases of divergence can reveal how differences in geographical or ecological context influence population histories and evolutionary trajectories. Here we examine a pair of sister species in the family Labridae: the yellowtail coris, *Coris gaimard* (Quoy & Gaimard, 1824), is distributed from Hawai'i and Polynesia to the eastern Indian Ocean (Randall, 1999; Fig. 1a). The African coris, *Coris cuvieri* (Bennett, 1831), inhabits the Red Sea and the Indian Ocean (Randall, 1999). These species are closely related, as indicated by overlapping counts of fin rays, lateral line scales and gill rakers, but differ in colouration (Randall, 1999, Fig. 1b, d).

These *Coris* wrasses are small (<20 cm), generalist predators that favour reefs and associated coral rubble (Randall, 1999; Ferry-Graham *et al*., 2002).The maximum PLD for *C. gaimard* is approximately 53 days (Victor, 1986). Other *Coris* species show similar PLDs (Victor, 1986) so it is here assumed that the PLD for *C. cuvieri* is approximately the same as that of *C. gaimard.* The juveniles of both species look strikingly different from their adult counterparts and appear to mimic clownfishes (*Amphripion* spp.) (Randall, 2005; Reininger, 2011, Fig. 1e, f, g). Clownfish obtain protection from predation by association with anemones. The *Coris* mimic, found in close association with clownfish habitat, is expected to benefit by reduced predation and possibly deception of competitors (Reininger, 2011). Notably, this mimic coloration is observed in Hawai'i where clownfishes do not occur. Batesian mimicry theory holds that an edible or harmless species (the mimic) resembles an inedible or harmful species (the model) in order to avoid predation (Caley & Schluter, 2003; Randall, 2005). In locations where the model is not present, the predators cannot learn to avoid the model, putting the mimic at higher risk of predation (Pfennig *et al*., 2001). Therefore Batesian mimicry should break down in areas where the model is not present, or the mimic could go extinct (Harper & Pfennig, 2008). Neither of these scenarios has occurred in *C. gaimard*, which invokes questions about how long the mimic has inhabited Hawai'i, and whether gene flow continues to connect Hawai'i to other Pacific populations with a potentially maladapted life stage.

In addition to investigating the origins of the Indian and Pacific *Coris* taxa, we are interested in the role of peripheral biogeographical provinces (Red Sea and Hawai'i), which

may serve as evolutionary incubators for the production of new species (Bowen *et al.,* 2013; DiBattista *et al.,* 2013). This range-wide survey of *C. gaimard* and *C. cuvieri* is therefore intended to evaluate evolutionary processes and address the following questions: A) Do these sister species share similar patterns of population structure across their ranges? B) Do genetic partitions coincide with known biogeographical barriers? C) When and how did these two species diverge? D) Could on-going gene flow explain the mimicry of juvenile *C. gaimard* in Hawai'i, given the absence of the model? This range wide study of two closely related sister species covers the entire Indo-Pacific and Red Sea and with this coverage we hope to provide insight about the phylogeography and evolution of Indo-Pacific reef fishes.

MATERIALS AND METHODS

Sample Collection and Extraction

Samples (*C. gaimard, n* = 337; *C. cuvieri, n* = 76, putative hybrids, *n* = 13) were collected from 20 locations (Fig. 1a) between 2002 and 2014. A twelve year sampling regime was necessary to obtain relatively complete coverage of a widely-distributed species pair. Fin clips or gill filaments were collected from each specimen and preserved in saturated salt solution (with 20% DMSO) or 95% ethanol, and stored at room temperature. Specimens were identified morphologically in the field by ichthyologists trained to recognise these species. DNA was extracted using the modified HotSHOT DNA extraction protocol (Meeker *et al*., 2007) or E-Z 96 Tissue DNA Kit (Omega, Norcross, Georgia, USA) following the manufacturer's instructions. As indicated in Tables 1 & 2, not all specimens were used in all analyses due to variation in PCR amplification success.

DNA Sequence Production

A 561 base pair (bp) segment of the mitochondrial DNA (mtDNA) cytochrome oxidase *c* subunit 1 (COI) was resolved using fish specific primers FishF2 and FishR2 (Ward *et al*., 2005). When these primers did not amplify, overlapping FISH BCL and FISH BCH primers (Baldwin *et al.,* 2009) were used instead, and all sequences were trimmed to the same length. To evaluate congruent phylogenetic relationships across multiple loci, primers GnRH3F and GnRH3R (Hassan *et al*., 2002) were used to resolve 286 bp of the third intron in the gonadotropin-releasing hormone (GnRH), and primers S7RPEX2R and S7RPEX2F (Chow

& Hazama, 1998) were used to resolve 542 bp of the second intron of the ribosomal protein S7. Details of the PCR reactions and DNA fragment preparation are provided in Supplemental Materials (Appendix S1) and are also available in Ahti (2015).

Data Analysis

DNA sequences were aligned, edited and trimmed to a common length using GENEIOUS 6.1.6 (Biomatters, LTD, Auckland, NZ) and analysed using DNASP 5.10 (Librado & Rozas, 2009). The phase of diploid nuclear sequences was reconstructed using PHASE (Stephens & Donnelly, 2003) as implemented in DNASP (Librado & Rozas, 2009), using 100,000 iterations and 10,000 burn-in iterations. Analyses of molecular variance (AMOVA), observed (H_o) and expected (H_E) heterozygosities and deviations from Hardy–Weinberg equilibrium were tested with ARLEQUIN 3.11 (Excoffier *et al*., 2005) using 1,000,000 Markov chain steps and 100,000 dememorisation steps.

JMODELTEST 2.1.4 (Guindon & Gascuel, 2003; Darriba *et al*., 2012) with Akaike Information Criterion (AIC) was employed to determine the best-fit model of DNA evolution for each data set. The Tamura & Nei (1993) model was selected as the overall best fit model. To evaluate genetic diversity, ARLEQUIN was used to calculate haplotype diversity (*h*) and nucleotide diversity (π) for each location. Population pairwise ϕ_{ST} (an analog of F_{ST} that includes sequence divergence) was calculated with 10,000 permutations using ARLEQUIN. False discovery rate (FDR) was implemented using the method of Benjamini & Hochberg (2009). Isolation by distance was assessed by plotting pairwise ϕ_{ST} against geographical distance (km) and using a Mantel Test with 9,999 permutations in GENALEX 6.5 (Peakall & Smouse, 2006, 2012). Fu's F_s test (Fu, 1997) with 1,000 simulated samples was carried out with ARLEQUIN to assess demographic history. This parameter measures excesses of low-frequency haplotypes, an indicator of selection or (more often) population expansion. The nucleotide divergence (*d*) between species and the overall Fu's F_s for both species was computed using DNASP.

The time since most recent population expansion was calculated for COI only as no suitable molecular clock exists for GnRH and S7. It was calculated as follows: $\tau = 2vt$, where τ for each population was obtained from ARLEQUIN, and v = mutation rate per lineage (3% per Myr between lineages for COI in wrasses; Ludt *et al.*, 2012) x (10^{-6} for COI) x the

sequence length. To transform t into years it was multiplied by the estimated generation time of 2 years (R.J. Toonen, pers. comm.). Although the generation time is only an approximation, we are largely interested in relative versus absolute estimates of expansion time. These dates should be regarded as first order approximations. To estimate the time to most recent common ancestor (TMRCA), the data was formatted using the programme BEAUTI 1.4.7 and a Bayesian Markov Chain Monte Carlo (MCMC) approach was implemented in BEAST 2.2.0 (Drummond & Rambaut, 2007). The analysis was conducted with a strict clock of 3% per million years between lineages (Ludt *et al.,* 2012) using a coalescent tree prior assuming exponential growth. *Coris formosa* (GenBank Accession No. KF929780), the closest known relative to our study species, was used as an outgroup. Default priors under the HKY+G+I model of mutation were used and simulations ran for 10 million generations with sampling every 1000 generations. Ten independent runs were compared to ensure convergence, and log files were combined and ages averaged across runs using TRACER 1.6 (http://tree.bio.ed.ac.uk/software/tracer/).

To infer relationships between populations and species, a haplotype network for each locus was created using NETWORK 4.6.1.2 (www.fluxus-engineering.com/network_terms.htm).

RESULTS

While there were diagnostic mtDNA differences between species, six specimens originally noted for intermediate colouration (Fig. 1c), indicative of hybrids, had a *C. gaimard* haplotype. Additionally, seven individuals with *C. gaimard* or intermediate colouration had a *C. cuvieri* haplotype. All of these 13 specimens were captured at Christmas Island in the eastern Indian Ocean, and have been excluded from population genetic analyses. After excluding putative hybrids, total sample size (*n*) was 76 for *C. cuvieri*, and 337 for *C. gaimard*. Locations where $n < 5$ were excluded from population genetic comparisons, but retained in haplotype networks.

Coris cuvieri

Mitochondrial DNA

A total of 73 COI sequences were obtained from *C. cuvieri*, revealing 9 polymorphic sites and 11 haplotypes (Table 1). The most common haplotype was observed in all locations (Fig. 2). Overall haplotype diversity was h = 0.269 and nucleotide diversity π = 0.0006. Population pairwise ϕ_{ST} (AMOVA) comparisons detected no population structure, with the exception of Djibouti compared to the Seychelles and Red Sea (ϕ_{ST} = 0.047 and 0.110, respectively, P < 0.05 in both cases) (Table S1). However, this differentiation was not significant when corrected for FDR. No isolation by distance was observed (r^2 = 0.073, P = 0.290). Population parameters for each location are provided in Table 1. The overall Fu's F_s for *C. cuvieri* was -12.349 (P < 0.001), indicating recent population expansion. The time since most recent population expansion was estimated to be approximately 38,000 to 178,000 years.

Nuclear DNA

When suspected hybrids were removed, 61 individuals were amplified and sequenced for GnRH and 51 for S7 (Table 2). The 286 bp segment of GnRH locus had seven polymorphic sites that yielded seven alleles (Fig. 3a), and the 542 bp segment of the S7 locus had 22 polymorphic sites that yielded 22 alleles (Fig. 3b). Pairwise ϕ_{ST} comparisons for GnRH showed differentiation between Djibouti and the Seychelles (ϕ_{ST} = 0.059, P = 0.024), but this was not significant after correcting for FDR (Table S1 in Appendix S2). The S7 locus showed no population genetic structure (Table S2 in Appendix S2). No isolation by distance was observed for either loci (r^2 = 0.297, P = 0.125 for GnRH; r^2 = 0.175, P = 0.202 for S7). Both nuclear markers were in Hardy–Weinberg equilibrium (Table 2).

Coris gaimard

Mitochondrial DNA

After excluding suspected hybrids (n = 7), the analysis of 300 individuals showed 26 polymorphic sites and 23 haplotypes for COI, with overall haplotype diversity h = 0.380 and nucleotide diversity π = 0.0009 (Table 1). The most common haplotype was found in each location (Fig. 2).

The Main Hawaiian Islands (MHI), Northwestern Hawaiian Isands (NWHI), and adjacent Johnston Atoll showed no genetic differentiation but were each significantly differentiated (ϕ_{ST} = 0.040 to 0.173, P < 0.01) from locations to the south and to the west: Kiribati,

Moorea, Palau and Christmas Island (Table 3). Structure was also observed between Moorea and Christmas Island (ϕ_{ST} = 0.130, P = 0.002), at opposite ends of the range, and Moorea and Kiribati (ϕ_{ST} = 0.100, P = 0.022), but the latter was not significant when correcting for FDR. No significant isolation by distance was observed (r^2 = 0.037, P = 0.790).

Population parameters for each location are summarised in Table 1. Fu's F_s for the overall data set was -31.227 (P < 0.001), indicating selection or (more likely) a recent population expansion. The time of most recent expansion for *C. gaimard* populations was 31,000 to 178,000 years. The older expansion dates were from the northern Pacific region; younger expansion dates were from the central and southern Pacific region and eastern Indian Ocean.

Nuclear DNA

After removing putative hybrids, a total of 286 bp of the GnRH intron was resolved from 278 specimens and 542 bp of the S7 intron from 225 specimens (Table 2). No S7 sequences were successfully amplified from the Philippines or Cook Islands. The GnRH locus had seven polymorphic sites defining five alleles (Fig 3a). The most common GnRH allele was observed in each sampled location for both species, excluding the Paracel Islands where the sample size was low (n = 2). The S7 locus had 24 polymorphic sites defining 34 alleles (Fig. 3b).

Pairwise ϕ_{ST} values for GnRH showed no significant population structure among *C. gaimard* samples when corrected for FDR. Using the standard P < 0.05 as a significance level, Christmas Island was differentiated compared to MHI, NWHI and Johnston Atoll (ϕ_{ST} = 0.045 to 0.088) (Table 4). No significant isolation by distance (r^2 = 0.086, P = 0.052) was detected. Johnston Atoll and the Cook Islands were out of Hardy–Weinberg equilibrium with an excess of homozygotes (P = 0.001 and 0.006, respectively) (Table 2). Removing these two locations from the analysis had no effect on population delineation and corresponding conclusions, so they were retained.

Pairwise ϕ_{ST} comparisons for S7 revealed patterns similar to those observed with COI. MHI, NWHI and Johnston Atoll were differentiated (ϕ_{ST} = 0.026 to 0.075, P < 0.05) from other locations: Moorea, Palau and Christmas Island (Table 3). Additionally, MHI, and Johnston Atoll were both differentiated from Kiribati (ϕ_{ST} = 0.033 and 0.057, respectively, P

< 0.05), and Johnston Atoll was differentiated from Cocos-Keeling (ϕ_{ST} = 0.072, *P* = 0.041). However, when the pairwise ϕ_{ST} comparisons were controlled for FDR, genetic structure was only detected between Johnston Atoll and Kiribati, Moorea, Palau and Christmas Island (ϕ_{ST} = 0.056 to 0.075, *P* < 0.005), as well as between the Main Hawaiian Islands and Christmas Island (ϕ_{ST} = 0.052, *P* < 0.001). Isolation by distance was not significant (r^2 = 0.059, *P* = 0.118). After correction for FDR, only Christmas Island was out of Hardy–Weinberg equilibrium for S7 (Table 2).

C. gaimard* versus *C. cuvieri

The nucleotide divergence between species was *d* = 0.0097 for COI (Fig. 2), with no diagnostic differences at the two nuclear loci (Fig 3a, 3b). Bayesian MCMC analysis determined the time to most recent common ancestor as 2.12 Ma (1.05 - 3.36 Ma) for *C. cuvieri* and 1.76 Ma (0.90 - 2.74 Ma) for *C. gaimard.*

Pairwise comparisons for COI between *C. gaimard* and *C. cuvieri* indicated strong population genetic differentiation between the oceans (ϕ_{ST} = 0.911, *P* < 0.001). No significant structure between oceans was observed with GnRH (ϕ_{ST} = 0.010, *P* = 0.064). However, significant structure in GnRH was detected when *C. gaimard* from the Hawaiian biogeographical province (MHI, NWHI, and Johnston) were compared to *C. gaimard* in the rest of the Pacific Ocean (ϕ_{ST} = 0.022, *P* < 0.005) and against *C. cuvieri* in the Indian Ocean (ϕ_{ST} = 0.035, *P* < 0.003). The structure between the Hawaiian Province and the rest of the Pacific (incl. Christmas Island) is probably an artefact of the structure between Christmas Island and the Hawaiian Islands as discussed above. For the S7 locus, significant population genetic structure was observed between species (ϕ_{ST} = 0.129, *P* < 0.001). When MHI, NWHI, and Johnston Atoll were compared against the rest of the Pacific Ocean and *C. cuvieri* in the Indian Ocean, genetic structure was observed between these groups (Hawai'i and Johnston Atoll versus Pacific, ϕ_{ST} = 0.046; Hawai'i and Johnston Atoll versus Indian Ocean, ϕ_{ST} = 0.162; Pacific versus Indian Ocean, ϕ_{ST} = 0.113, *P* < 0.001).

DISCUSSION

The Indo-Pacific includes a biota that uniquely spans more than two-thirds of the planet. The centre of this range hosts both the highest marine biodiversity, and the largest

biogeographical region on Earth, the Indo-Polynesian Province that extends from the central Pacific to the western Indian Ocean (Briggs & Bowen, 2012). On the periphery of this vast area are biogeographical provinces including the Red Sea, Western Indian Ocean and Hawaiian Archipelago, distinguished by high levels of endemism and (more recently) genetic partitions within species (DiBattista *et al*., 2011, 2013 for example). Our study group, encompassing the sister species *C. cuvieri* and *C. gaimard*, inhabits this entire region and can provide insight into population connectivity and the earliest stages of evolutionary divergence in the marine environment.

In the Indian Ocean, pairwise ϕ_{ST} comparisons revealed no population structure for *C. cuvieri* following FDR correction. The lack of significant structure between the Indian Ocean and Red Sea indicates high connectivity, a surprising result given that the sampling locations are spread across more than 8000 km with vast discontinuities between reef habitats. It may be possible that long PLD (53 days in the sister *C. gaimard*) enables larvae to traverse stretches of ocean and biogeographical barriers in this region. Another likely factor is that these species are capable of living in degraded habitat, and therefore more likely to successfully colonize, recruit and survive to reproduction in a wide array of reefs. Other life history traits (not only PLD) may therefore play a role in connectivity (Keith *et al*., 2015). This finding contrasts with a survey of the congeneric *Coris julis* in the Mediterranean and NW Atlantic, which reported significant structure between ocean basins (Fruciano *et al*., 2011). In the Pacific Ocean, the only consistent structure detected in *C. gaimard* pairwise ϕ_{ST} comparisons was the isolation of Hawaii and Johnston Atoll from other central Pacific locations.

Red Sea connectivity

The Strait of Bab al Mandab is the only connection between the Red Sea and the Indian Ocean. During glacial sea level fluctuations, the closure of this strait caused extreme changes in salinity and temperature inside the Red Sea (Siddal *et al*., 2003). The southern Red Sea is also characterised by eutrophic conditions and sparse reef habitat, possibly acting as a contemporary barrier to dispersal (Roberts *et al*., 1992). Some reef fishes can traverse these barriers, while others show deep intraspecific partitions between the Red Sea and the Indian Ocean (DiBattista *et al*., 2013; Fernandez-Silva *et al*., 2015). The lack of population

318 genetic structure between these two ocean basins may indicate that *C. cuvieri* in the Red
319 Sea is a recent arrival, or is capable of maintaining contemporary gene flow with the Indian
320 Ocean.

321 Under the conventional $P < 0.05$ significance level (in contrast to the FDR),
322 population structure was detected between Djibouti and the Seychelles in COI and GnRH,
323 and between Djibouti and the Red Sea in COI. The Seychelles are regarded as part of the
324 Western Indian Ocean biogeographic province (Briggs & Bowen, 2012), and further sampling
325 could illuminate this connection between the eastern coast of Africa and the Red Sea, to the
326 exclusion of the intermediate Djibouti. The signal of structure between Djibouti and the Red
327 Sea in addition to the Seychelles is particularly surprising given that Djibouti is located on
328 the only natural gateway between the Indian Ocean and the Red Sea. Contemporary
329 barriers are unlikely to have caused this rift given the wide dispersal and high connectivity of
330 this *Coris* species. Interestingly, the times of expansion for the Red Sea and Seychelles are
331 much more recent (38 to 48 kyr) than for Djibouti and Diego Garcia (178 Ka). This may be a
332 relic effect of habitat alterations caused by glacial sea level changes. The Red Sea and the
333 Seychelles may also be more vulnerable to habitat alterations than Djibouti due to the
334 topography and geography of these regions (see below) (Obura, 2012).

335 **Indo-Pacific population history**

336 Significant negative Fu's F_S values in *C. gaimard* in Hawai'i, Johnston Atoll and Christmas
337 Island indicate either selection or (more likely) population expansion after a bottleneck or
338 founding event. Under a conventional molecular clock, these demographic events are dated
339 to approximately 178 Ka in the northern Pacific. In contrast, the southern and western
340 Pacific as well as Christmas Island show more recent population expansion approximately 30
341 to 40 Ka. While these absolute estimates are not precise, they indicate a strong difference in
342 the timing of events.

343 The two geographically isolated provinces at the ends of the range (Hawai'i and the
344 Red Sea) show contrasting population histories. In the Pacific the oldest mtDNA expansion
345 time is observed in the Hawaiian Archipelago and Johnston Atoll, while in the Indian Ocean
346 the isolated Red Sea has the most recent expansion. Estimates of expansion time and
347 negative Fu's F_s values for *C. cuvieri* indicate population expansion approximately 38 to 48

Ka in the Red Sea and the Seychelles. As the most recent expansion event was detected in the Red Sea, this could indicate that the population did not persist through the hypersaline conditions associated with Pleistocene glaciations, as has been suggested for multiple marine species (DiBattista *et al*., 2013). However, due to the low sample size (*n* = 11) from the Red Sea, these results should be interpreted with caution. The younger population in the Seychelles may be due to the relatively shallow (< 200 m) Mascarene Plateau that underlies this archipelago (New *et al*., 2007). Lower sea level during glaciation may have reduced reef habitat in this area, causing extirpation or population reduction as in the Red Sea. Older populations in Diego Garcia (178 Ka) could be explained by the deep topography of the Chagos Trench that may have provided refugia for reef fishes during sea level fluctuations (Ludt & Rocha, 2015). Similar expansion dates in Djibouti could be due to a possible refuge in the Gulf of Aden or Arabian Sea (Klausewitz, 1989). Notably, the locations with older coalescence times (Hawai'i, Diego Garcia, and Djibouti), are all near deep ocean trenches. These locations may have provided refugia for reef fishes during the glacial sea level fluctuations, although these patterns are not universal (DiBattista *et al.,* 2015).

Shallow divergence between *C. gaimard* and *C. cuvieri* and the discovery of hybrids

Pleistocene sea levels, 120m to 140m below present, created a nearly-complete separation of Indian and Pacific marine fauna, leading to intraspecific genetic divergence in many taxa, and ultimately the formation of sister species pairs (McMillan *et al*., 1999; Barber *et al*., 2006; Gaither & Rocha, 2013; Ludt & Rocha, 2015). The close relationship between these two *Coris* species is illustrated by shared alleles for both GnRH and S7 (Fig. 3a, b), but not with COI haplotypes (Fig. 2). This finding is similar to the parrotfish *Chlorurus sordidus* (family Labridae), which shows two mtDNA lineages (*d* = 0.071 in mtDNA control region) but no sharing of haplotypes between Indian and Pacific Ocean cohorts (Bay *et al.,* 2004). The grouper *Cephalopholus argus* also has Indian and Pacific Ocean lineages separated by *d* = 0.008 in mtDNA cytochrome *b*, with a low level of overlap between lineages (Gaither *et al*., 2011). The COI divergence of *d* = 0.0097 between *C. cuvieri* and *C. gaimard* in the Indian and Pacific Oceans corresponds to about 2 Ma, according to the Bayesian approach. However, divergences of *d* < 0.01 between sister species are uncommon. Grant & Bowen (1998) summarised mtDNA divergences in marine sister species pairs, showing partitions of *d* = 0.034 to 0.124 based on three different markers (cytochrome *b*, ND4/5, and COI),

concordant with other surveys of vertebrate sister species (Johns & Avise, 1998; Guillemaud *et al*., 2000). Within-species comparisons of Brazilian and Caribbean wrasses revealed mtDNA cytochrome *b* divergences of *d* = 0.023 (*Halichoeres cyanocephalus*) to *d* = 0.065 (*Halichoeres maculipinna*) (Rocha, 2004). Atlantic sister species *Coris julis* and *C. atlantica* are distinguished by *d* = 0.045 in mitochondrial 12S rDNA (Guillemaud *et al*., 2000). Hence the observed divergence between *C. gaimard* and *C. cuvieri* falls well within the range of intraspecific partitions in other reef fishes, including wrasses. Additional studies may reveal that *C. gaimard* and *C. cuvieri* are more appropriately regarded as subspecies.

Closely related reef fishes that have arisen by allopatric speciation are likely to hybridise upon secondary contact (Montanari *et al*., 2014). Christmas Island, adjacent to the Indian-Pacific Ocean border, is a hybridisation hotspot for shallow marine fauna (Hobbs *et al*., 2009). The presence of *C. gaimard* x *C. cuvieri* hybrids at Christmas Island shows that the genetic and ecological factors that facilitated speciation are not sufficient to impose complete reproductive isolation. Although *C. cuvieri* has not been recorded at Christmas Island (Hobbs *et al.*, 2014), the finding of *C. cuvieri* haplotypes in individuals with *C. gaimard* colouration (or that appeared intermediate between the two species; Fig. 1c) indicates that hybrid larvae may be produced elsewhere and reach Christmas Island by pelagic dispersal. A similar situation was observed with butterflyfish hybrids (*Hemitaurichtys zoster* x *H. polylepsis*) at Christmas Island, where parental species *H. zoster* has not been recorded (Hobbs & Allen, 2014). Given that *C. cuvieri* x *C. gaimard* hybrids have also been observed in Indonesia (Bali, Tulamben; R.F. Myers, pers. comm.) it is likely that range overlap and hybridisation occur elsewhere along the Indian-Pacific boundary.

The Hawaiian Archipelago

No structure was observed between the three regions within the Hawaiian Province (MHI, NWHI, and Johnston Atoll), a common finding in reef fishes (Toonen *et al*., 2011; Selkoe *et al*., 2014). Johnston Atoll is located approximately 1400 km south-west of Hawai'i, and is regarded as part of the Hawaiian biogeographical province based on faunal similarity (Briggs & Bowen, 2012). Simulated larval transport models indicate a potential dispersal corridor between Johnston Atoll and Hawai'i (Kobayashi, 2006) and genetic connectivity studies support this connection (Leray *et al*., 2010; Andrews *et al*., 2014). Species with

relatively long PLD (similar to *Coris*) show genetic connectivity between Hawai'i and Johnston (Eble *et al*., 2011), while others do not (DiBattista *et al*., 2011).

The strongest signal of population structure in either species is between the Hawaiian Province and other Pacific locations, as indicated by COI and S7 markers (Table 3). In this study, Hawai'i is significantly isolated from Kiribati, Moorea, Palau and Christmas Island but not from Cocos-Keeling, Cook Islands and the Philippines, possibly an artefact of low sample sizes ($n \leq 13$) in the latter locations. However, the lack of structure between Hawai'i and the Philippines, along with similar estimates of time since expansion (177 to 178 kyr), may indicate contemporary connectivity in the northern hemisphere, promoted by the Kuroshio and North Pacific Currents.

The Hawaiian Mimic

The clownfish mimicry by *C. gaimard* in Hawai'i is an evolutionary enigma. If the model is absent, then bright colouration should invoke a heavy cost in predation pressure. The mimicry has not broken down nor has the mimic gone extinct, thus contradicting Batesian mimicry theory.

Harper & Pfennig (2008) suggested that male-mediated dispersal could explain the presence of mimicry in allopatry. However, they showed that mimicry is less accurate in allopatry than it is in sympatry. The wrasse *C. gaimard,* like most reef fishes, is a broadcast spawner and therefore dispersal occurs before gender-specific traits emerge. Further, the isolation of Hawai'i shows up in both maternal (mtDNA) markers and biparentally-inherited nuclear markers. Thus, male-mediated gene flow cannot explain the presence of mimicry in Hawaiian *C. gaimard*.

The presence of *C. gaimard* clownfish mimics in Hawai'i and Johnston Atoll should indicate that it is either a recent arrival to the region or that sufficient gene flow between Hawai'i and other locations maintains the mimic colouration. Given that *C. gaimard* has a long history in Hawai'i (estimated at least 178 kyr) and the mimicry has not yet broken down, perhaps there is sufficient gene flow across the Northern Pacific (Philippines to Hawai'i) to maintain the mimetic pattern in this species, despite the isolation of Hawai'i from most of the Pacific.

Conclusion

Here we conducted a range-wide genetic survey of sister taxa *C. gaimard* and *C. cuvieri* to address four questions:

A) *Do these sister species share similar patterns of population structure across their ranges?*

Both species show high dispersal ability, with no significant population structure across thousands of kilometres. The Indian Ocean lineage shows no population distinction of the Red Sea, whereas the Pacific lineage shows a strong isolation of the Hawaiian Archipelago from other central Pacific locations, but also evidence of connectivity across the North Pacific between Philippines and Hawaii.

B) *Do the observed genetic patterns coincide with known biogeographical barriers*?

The taxonomic and genetic split between *C. cuvieri* and *C. gaimard* is coincident with the intermittent Indo-Pacific Barrier. The population in the Hawaiian Biogeographical Province is isolated, but the cohort in the Red Sea Province is not.

C) *When and how did these two species diverge?*

Species divergence estimates range from 0.5 Ma, based on a conventional molecular clock, to approximately 2 Ma, based on Bayesian analysis. In either case, time estimates lie within the Pleistocene glacial cycles and indicate that the primary evolutionary partition is at the Indo-Pacific Barrier. Shallow morphological and genetic divergences also indicate that these sister species may be more appropriately recognised as subspecies.

D) *Could on-going gene flow explain the clownfish mimicry of juvenile* C. gaimard *in Hawai'i?*

We found a relatively old estimate of population expansion in Hawai'i (178 kyr), in addition to isolation from other sample locations in the central Pacific. However, the Hawaiian population was not significantly different from the Philippines in population genetic comparisons, invoking the possibility of connections through the North Pacific gyre. The mimicry in Hawaii may be explained by ongoing gene flow. Overall, the history we resolved in *Coris* wrasses appears to be a balance between high dispersal ability and

recruitment success, which homogenises populations even with costly mimicry traits, versus the geographic and oceanographic isolation that may promote evolutionary divergence.

ACKNOWLEDGEMENTS

For sample collections and logistic support we thank Howard Choat, Pat Collins, Laurie Collins, Matthew Craig, Toby Daly-Engel, Matt Dunlap, Jeff Eble, Iria Fernandez-Silva, Michelle Gaither, Brian Greene, Jean-Paul Hobbs, Matt Iacchei, Stephen Karl, Randall Kosaki, Jo-Ann Leong, Eric Mason at Dream Divers, Carl Meyer, Gerrit Nanninga, Yannis Papastamatiou, David Pence, Daniel Polhemus, Nicolas Prévot at Dolphin Divers and the crew of the M/V Deli, Zeng Xiaoqi from the Ocean University of China College of Fisheries and Song He, Mark Priest, Joshua Reece, Matt Ross, Jennifer Schultz, Tane Sinclair-Taylor, Zoltan Szabo, Robert Toonen, Jacquelyn Troller, Jill Zamzow, the KAUST Coastal and Marine Resources Core Lab and Amr Gusti, the crew of the R/V Thuwal at KAUST, the crew of the R/V Hi'ialakai, Coral Reef Research Foundation, Papahanaumokuakea Marine National Monument, NOAA National Marine Sanctuaries, U.S. Fish and Wildlife Service, and the State of Hawaii Division of Aquatic Resources, the University of Milano-Bicocca Marine Research and High Education Centre in Magoodhoo, the Ministry of Fisheries and Agriculture, Republic of Maldives, and the community of Maghoodhoo, Faafu Atoll, Western Australia Department of Fisheries, Parks Australia, and everybody involved in the logistics of the expeditions. This work was supported by U.S. National Science Foundation grants OCE09-29031 to BWB, and OCE12-60169 to R.J. Toonen, the Seaver Institute (BWB), NOAA National Marine Sanctuaries Program MOA 2005 – 008/6682 (R.J. Toonen), the KAUST Office of Competitive Research Funds (OCRF) under Award No. CRG-1-2012-BER-002 and baseline research funds to M.L.B.,a National Geographic Society Grant 9024-11 to J.D.D., NSF grant DGE-1329626 to RRC and Dr. Nancy Foster Scholarship program under Award no. NA15NOS4290067 (RRC). Thanks to associate editor Michelle Gaither and two anonymous reviewers who improved the manuscript.

REFERENCES

Ahti, P.A. (2015) Phylogeography of *Coris gaimard* and *Coris cuvieri* in the Pacific Ocean, Indian Ocean, and Red Sea. MSci. thesis, University of Glasgow, Scotland.

493 Andrews K.R., Moriwake, V.N., Wilcox, C., Grau, E.G., Kelley, C., Pyle, R.L. & Bowen, B.W.
494 (2014) Phylogeographic analyses of submesophotic snappers *Etelis coruscans* and *Etelis*
495 "*marshi*" (Family Lutjanidae) reveal concordant genetic structure across the Hawaiian
496 Archipelago. *PloS One*, **9**, e91665.

497 Barber, P.H., Erdmann, M.V. & Palumbi, S.R. (2006) Comparative phylogeography of three
498 codistributed stomatopods: origins and timing of regional lineage diversification in the
499 Coral Triangle. *Evolution*, **60**, 1825–1839.

500 Baldwin, C.C., Mounts, J.H., Smith, D.G. & Weight, L.A. (2009) Genetic identification and
501 color descriptions of early life-history stages of Belizean *Phaeoptyx* and *Astrapogon*
502 (Teleostei: Apogonidae) with comments on identification of adult *Phaeoptyx*. *Zootaxa*,
503 **22**, 1-22.

504 Bay L.K., Choat, J.H., van Herwerden, L. & Robertson, D.R. (2004) High genetic diversities and
505 complex genetic structure in an Indo-Pacific tropical reef fish (*Chlorurus sordidus*):
506 evidence of an unstable evolutionary past? *Marine Biology*, **144**, 757–767.

507 Benjamini, Y. & Hochberg, Y. (2009) Controlling the false discovery rate: a practical and
508 powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*
509 *(Methodological)*, **57**, 289–300.

510 Bird, C.E., Holland, B.S., Bowen, B.W. & Toonen, R.J. (2011) Diversification of endemic
511 sympatric limpets (*Cellana* spp.) in the Hawaiian Archipelago. *Molecular Ecology*, **20**,
512 2128 – 2141.

513 Bowen, B.W., Rocha, L.A., Toonen, R.J., Karl, S.A., Craig, M.T., DiBattista, J.D, Eble, J.A.,
514 Gaither, M.R., Skillings, D. & Bird, C.E. (2013) The origins of tropical marine biodiversity.
515 *Trends in Ecology & Evolution*, **28**, 359–366.

516 Briggs, J.C. & Bowen, B.W. (2012) A realignment of marine biogeographic provinces with
517 particular reference to fish distributions. *Journal of Biogeography*, **39**, 12–30.

518 Briggs, J.C. & Bowen, B.W. (2013) Evolutionary patterns: Marine shelf habitat. *Journal of*
519 *Biogeography*, **40**, 1023 – 1035.

Caley, M.J. & Schluter, D. (2003) Predators favour mimicry in a tropical reef fish. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 667–672.

Chow, S. & Hazama, K. (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*, **7**, 1255–1256.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

DiBattista, J.D., Wilcox, C., Craig, M.T., Rocha, L.A. & Bowen, B.W. (2011) Phylogeography of the Pacific blueline surgeonfish, *Acanthurus nigroris*, reveals high genetic connectivity and a cryptic endemic species in the Hawaiian Archipelago. *Journal of Marine Biology*, Article ID 839134.

DiBattista, J.D., Rocha, L.A., Craig, M.T., Feldheim, K.A. & Bowen, B.W. (2012) Phylogeography of two closely related Indo-Pacific butterflyfishes reveals divergent evolutionary histories and discordant results from mtDNA and microsatellites. *Journal of Heredity*, **103**, 617–629.

DiBattista, J.D., Berumen, M.L., Gaither, M.R., Rocha, L.A., Eble, J.A, Choat, J.H., Craig, M.T., Skillings, D.J. & Bowen, B.W. (2013) After continents divide: comparative phylogeography of reef fishes from the Red Sea and Indian Ocean. *Journal of Biogeography*, **40**, 1170–1181.

DiBattista, J.D., Roberts, M., Baird, A.H., Benzoni, F., Bouwmeester, J., Bowen, B.W., Coker, D.J., Lozano-Cortes, D.F., Choat, J.H., Gaither, M.R., Hobbs, J.-P.A., Khalil, M.T., Kochzius, M., Myers, R., Paulay, G., Robitzch, V.S.N., Saenz-Agudelo, P., Salas, E., Sinclair-Taylor, T.H., Toonen, R.J., Westneat, M.W., Williams, S.T. & Berumen, M.L. (2015) A review of contemporary patterns of endemism for shallow water reef fauna in the Red Sea. *Journal of Biogeography.* Online early.

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology,* **7**, 214.

546 Eble, J.A., Rocha, L.A., Craig, M.T. & Bowen, B.W. (2011) Not all larvae stay close to home:
547 Insights into marine population connectivity with a focus on the Brown Surgeonfish
548 (*Acanthurus nigrofuscus*). *Journal of Marine Biology*, Article ID 518516.

549 Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): An integrated software
550 package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**,
551 47–50.

552 Faurby, S. & Barber, P.H. (2012) Theoretical limits to the correlation between pelagic larval
553 duration and population genetic structure. *Molecular Ecology*, **21**, 3419–3432.

554 Fruciano, C., Hanel, R., Debes, P.V., Tigano, C., Ferrito, V. (2011) Atlantic-Mediterranean and
555 within-Mediterranean molecular variation in *Coris julis* (L. 1758) (Teleostei, Labridae).
556 *Marine Biology*, **158**, 1271-1286.

557 Fernandez-Silva, I., Randall, J.E., Coleman, R.R., DiBattista, J.D., Rocha, L.A., Reimer, J.D.,
558 Meyer, C.G., Bowen, B.W. (2015) Yellow tails in a Red Sea: Phylogeography of the Indo-
559 Pacific goatfish *Mulloidichthys flavolineatus* reveals isolation in peripheral provinces
560 and cryptic evolutionary lineages. *Journal of Biogeography. Online early.*

561 Ferry-Graham, L.A., Wainright, P.C., Westneat, M.W. & Bellwood, D.R. (2002) Mechanisms
562 of benthic prey capture in wrasses (Labridae). *Marine Biology*, **141**, 819 – 830.

563 Fu, Y-X. (1997) Statistical tests of neutrality of mutations against population growth,
564 hitchhiking and background selection. *Genetics*, **147**, 915–925.

565 Gaither, M.R. & Rocha, L.A. (2013) Origins of species richness in the Indo-Malay-Philippine
566 biodiversity hotspot: evidence for the centre of overlap hypothesis. *Journal of*
567 *Biogeography*, **40**, 1638–1648.

568 Gaither, M.R., Bowen, B.W., Bordenave, T.-R., Rocha, L.A., Newman, S.J., Gomez, J.A, van
569 Herwerden, L. & Craig, M.T. (2011) Phylogeography of the reef fish *Cephalopholis argus*
570 (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific barrier with
571 contemporary overlap in the coral triangle. *Evolutionary Biology*, **189**, 1–15.

572 Grant, W.S. & Bowen, B.W. (1998) Shallow population histories in deep evolutionary
573 lineages of marine fishes: Insights from sardines and anchovies and lessons for
574 conservation. *The Journal of Heredity*, **89**, 415–426.

575 Guillemaud, T., Cancela, M.L., Afonso, P., Morato, T., Santos, R.S. & Wirtz, P. (2000)
576 Molecular insights into the taxonomic status of *Coris atlantica* (Pisces: Labridae).
577 *Journal of the Marine Biological Association of the UK*, **80**, 929–933.

578 Guindon, S. & Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large
579 phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

580 Harper, G.R. & Pfennig, D.W. (2008) Selection overrides gene flow to break down
581 maladaptive mimicry. *Nature*, **451**, 1103–1106.

582 Hassan M., Lemaire C., Fauvelot, C. & Bonhomme, F. (2002) Seventeen new exon-primed
583 intron-crossing polymerase chain reaction amplifiable introns in fish. *Molecular Ecology*
584 *Notes*, **2**, 334–340.

585 Hobbs, J-P.A. & Allen, G.R. (2014) Hybridisation among coral reef fishes at Christmas Island
586 and the Cocos (Keeling) Islands. *Raffles Bulletin of Zoology,* **30**: 220-226.

587 Hobbs, J.-P.A., Frisch, A.J., Allen, G.R. & van Herwerden, L. (2009) Marine hybrid hotspot at
588 Indo-Pacific biogeographic border. *Biology Letters*, **5**, 258–261.

589 Hobbs, J-P.A., Newman, S.J., Mitsopoulos, G.E.A., Travers, M.J., Skepper, C.L., Gilligan, J.J.,
590 Allen, G.R., Choat, H.J. & Ayling, A.M. (2014) Checklist and new records of Christmas
591 Island fishes: the influence of isolation, biogeography and habitat availability on species
592 abundance and community composition. *Raffles Bulletin of Zoology,* **30**: 182-202.

593 Hourigan, T.F. & Reese, E.S. (1987) Mid-ocean isolation and the evolution of Hawaiian reef
594 fishes. *Trends in Ecology and Evolution*, **2**, 187–191.

595 Johns, G.C. & Avise, J.C. (1998) A comparative summary of genetic distances in the
596 vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and*
597 *Evolution*, **15**, 1481–1490.

598 Keith, S.A., Woolsey, P.S., Madin, J.S., Byrne, M. & Baird, A. (2015) Differential establishment
599 potential of species predicts a shift in coral assemblage structure across a biogeographic
600 barrier. *Ecography*, **38**, 1-10.

601 Kemp, J. (1998) Zoogeography of the coral reef fishes of the Socotra Archipelago. *Journal of*
602 *Biogeography*, **25**, 919–933.

603 Klausewitz, W. (1989) Evolutionary history and zoogeography of the Red Sea ichtyofauna.
604 *Fauna of Saudi Arabia*, **10**, 310-337.

605 Kobayashi, D.R. (2006) Colonization of the Hawaiian Archipelago via Johnston Atoll: a
606 characterization of oceanographic transport corridors for pelagic larvae using computer
607 simulation. *Coral Reefs*, **25**, 407–417.

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

611 Librado, P. & Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA
612 polymorphism data. *Bioinformatics*, **25**, 1451–1452.

613 Liu, S-Y.V., Chang, F-T., Borsa, P., Chen, W-J. & Dai, C-F. (2014) Phylogeography of the
614 humbug damselfish, Dascyllus aruanus (Linnaeus, 1758): evidence of Indo-Pacific
615 vicariance and genetic differentiation in peripheral populations. *Biological Journal of*
616 *the Linnean Society*, **113**, 931 – 942.

617 Ludt, W.B., Bernal, M.A., Bowen, B.W. & Rocha, L.A. (2012) Living in the past:
618 phylogeography and population histories of Indo-Pacific wrasses (genus *Halichoeres*) in
619 shallow lagoons versus outer reef slopes. *PloS one*, **7**, e38042.

620 Ludt, W.B. & Rocha, L.A. (2015) Shifting seas: the impacts of Pleistocene sea-level
621 fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, **42**, 25 –
622 38.

McMillan, W.O., Weigt, L.A. & Palumbi, S.R. (1999) Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). *Evolution*, **53**, 247–260.

Meeker, N.D., Hutchinson, S.A., Ho, L. & Trede, N.S. (2007) Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *BioTechniques*, **43**, 610–614.

Montanari, S.R., Hobbs, J-P.A, Pratchett, M.S., Bay, L.K. & van Herwerden, L. (2014) Does genetic distance between parental species influence outcomes of hybridization among coral reef butterflyfishes? *Molecular Ecology,* **11**, 2757-2770.

New, A.L., Alderson, S.G., Smeed, D.A. & Stansfield, K.L. (2007) On the circulation of water masses across the Mascarene Plateau in the South Indian Ocean. *Deep-Sea Research*, **54**, 42–74.

Obura, D. (2012) The diversity and biogeography of Western Indian Ocean reef-building corals. *PloS One*, **7**, e45013.

Peakall, R. & Smouse, P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288-295.

Peakall, R. & Smouse, P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, **28**, 2537–2539.

Pfennig, D.W., Harcombe, W.R. & Pfennig, K.S. (2001) Frequency-dependent Batesian mimicry. *Nature*, **410**, 323.

Randall, J.E. (1998) Zoogeography of shore fishes of the Indo-Pacific region. *Zoological Studies*, **37**, 227–268.

Randall, J.E. (1999) Revision of the Indo-Pacific fishes of the genus *Coris*, with descriptions of five new species*. Indo-Pacific Fishes,* **29,** 1-74

Randall, J.E. (2005) A review of mimicry in marine fishes. *Zoological Studies*, **44**, 299–328.

647 Randall, J.E. (2007) *Reef and Shore Fishes of the Hawaiian Islands*. University of Hawaii Sea
648 Grant Program, Honolulu, Hawaii.

649 Reece, J.S., Bowen, B.W., Smith, D. & Larson, A. (2011) Comparative phylogeography of four
650 Indo-Pacific moray eel species (Muraenidae) reveals comparable ocean-wide genetic
651 connectivity despite five-fold differences in available adult habitat. *Marine Ecology*
652 *Progress Series*, **437**, 269–277.

653 Reininger, M. (2011) Mimicry in juvenile wrasses: ecological and behavioural aspects of a
654 *Coris-Amphiprion* relationship in the northern Red Sea. *Zoology in the Middle East*, **54**,
655 23–34.

656 Roberts, C.M., Shepherd, A.R.D. & Ormond, R.F.G. (1992) Large-scale variation in
657 assemblage structure of Red Sea butterflyfishes and angelfishes. *Journal of*
658 *Biogeography*, **19,** 239–250.

659 Rocha, L.A. (2004) Mitochondrial DNA and color pattern variation in three western Atlantic
660 *Halichoeres* (Labridae), with the revalidation of two species. *Copeia*, **2004**, 770–782.

661 Rocha, L.A., Bass, A.L., Robertson, D.R. & Bowen, B.W. (2002) Adult habitat preferences,
662 larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes
663 (Teleostei: Acanthuridae). *Molecular Ecology*, **11**, 243–252.

664 Sale, P.F. (1980) The ecology of fishes on coral reefs. *Oceanography and Marine Biology*, **18**,
665 367 – 421.

666 Selkoe, K.A., Gaggiotti, O.E., ToBo Laboratory, Bowen, B.W. & Toonen, R.J. (2014) Emergent
667 patterns of population genetic structure for a coral reef community. *Molecular Ecology*,
668 **23**, 3064–3079.

669 Siddal, M., Rohling, E.J., Almongi-Labin, A., Hemleben, C., Meischner, D., Scmelzer, I. &
670 Smeed, D.A. (2003) Sea-level fluctuations during the last glacial cycle. *Nature*, **423**,
671 853–858.

672 Stephens, M. & Donnelly, P. (2003) A comparison of Bayesian methods for haplotype
673 reconstruction from population genotype data. *American Journal of Human Genetics*,
674 **73**, 1162–1169.

675 Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the
676 control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology*
677 *and Evolution*, **10**, 512–526.

678 Timm, J. & Kochzius, M. (2008) Geological history and oceanography of the Indo-Malay
679 Archipelago shape the genetic population structure in the false clown anemonefish
680 (*Amphiprion ocellaris*). *Molecular Ecology*, 17, 3999 – 4014.

681 Toonen, R.J., Andrews, K.R., Baums, I.B., Bird, C.E., Concepcion, G.T., Daly-Engel, T.S., Eble,
682 J.A., Faucci, A., Gaither, M.R., Iacchei, M., Puritz, J.B., Schultz, J.K., Skillings, D.J.,
683 Timmers, M.A. & Bowen, B.W.. (2011) Defining boundaries for ecosystem-based
684 management: A multispecies case study of marine connectivity across the Hawaiian
685 Archipelago. *Journal of Marine Biology*, **2011**, Article ID 460173.

686 Victor, B.C. (1986) Duration of the planktonic larval stage of one hundred species of Pacific
687 and Atlantic wrasses (family Labridae). *Marine Biology*, **90**, 317–326.

688 Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R. & Hebert, P.D.N. (2005) DNA barcoding
689 Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series*
690 *B, Biological sciences*, **360**, 1847–57.

691 **SUPPORTING INFORMATION**

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

693 **Appendix 1** PCR Reaction Conditions and DNA Fragment Preparation

694 Mitochondrial DNA.

695 **Appendix 2** Pairwise ϕ_{ST} comparisons for *Coris cuvieri*.

696 **BIOSKETCH**

The authors are focused on illuminating the evolutionary processes that generate marine biodiversity. They have carried out phylogeographical surveys of over 20 reef fish species in the Red Sea, Arabian Sea, and greater Indo-Pacific to test existing evolutionary models, to resolve the life history traits that influence dispersal and population separations in reef organisms and to inform marine conservation (e.g. defining the boundaries of marine protected areas).

Author contributions: PAA and RRC produced DNA sequences, analysed the data and collected tissue samples. PAA led the writing and RRC contributed to writing. BWB designed the study and procured funding. BWB, JDD, LAR, and MLB collected tissue samples and contributed to writing.

Table 1 Cytochrome *c* oxidase subunit I (COI) molecular diversity indices for *Coris cuvieri* and *Coris gaimard* populations. Sampling location, number of specimens (*n*), haplotype number (N_h), haplotype diversity (*h*), nucleotide diversity (π), Fu's F_s and associated *P*-value (significant at $P = 0.02$; Fu, 1997), τ (tau), and time since the most recent expansion (yrs). Infinity sign indicates value could not be resolved. South Africa ($n = 1$) and Maldives ($n = 1$) are only included in the total values.

Location	*n*	N_h	*h*	π	*Fu's* F_s	*P*	τ	Expansion
C. cuvieri								
Red Sea	11	3	0.473 ± 0.162	0.0009 ± 0.0009	-0.659	0.113	0.645	38,000
Djibouti	22	2	0.091 ± 0.081	0.0002 ± 0.0003	-0.957	0.007	3.000	178,000
Seychelles	15	5	0.562 ± 0.143	0.0012 ± 0.0011	-2.677	0.003	0.811	48,000
Diego Garcia	23	3	0.170 ± 0.103	0.0005 ± 0.0006	-1.305	0.035	3.000	178,000
Total	73	11	0.269 + 0.068	0.0006 + 0.0006	-12.349	<0.001	3.000	178,000
C. gaimard								
MHI	68	6	0.295 ± 0.071	0.0006 ± 0.0006	-4.626	0.001	3.000	178,000
NWHI	39	5	0.325 ± 0.094	0.0006 ± 0.0007	-3.289	0.005	2.980	177,000

Johnston	38	5	0.293 ± 0.096	0.0006 ± 0.0007	-5.378	0.000	3.000	178,000
Marshall Is.	9	2	0.389 ± 0.164	0.0007 ± 0.0008	0.477	0.408	0.543	32,000
Kiribati	25	4	0.530 ± 0.086	0.0013 ± 0.0011	-0.477	0.295	0.715	42,000
Moorea	16	4	0.517 ± 0.132	0.0024 ± 0.0018	0.426	0.565	0.000	∞
Cook Is.	9	2	0.222 ± 0.166	0.0004 ± 0.0006	-0.263	0.178	0.000	∞
Palau	21	5	0.486 ± 0.124	0.0015 ± 0.0012	-1.539	0.088	0.635	38,000
Philippines	13	2	0.282 ± 0.142	0.0005 ± 0.0006	0.240	0.325	2.982	177,000
Okinawa	5	4	0.900 ± 0.161	0.0040 ± 0.0031	-0.848	0.159	2.328	138,000
Christmas Is.	47	7	0.430 ± 0.088	0.0010 ± 0.0009	-4.418	0.003	0.523	31,000
Cocos Is.	8	2	0.250 ± 0.180	0.0004 ± 0.0006	-0.182	0.230	2.930	174,000
Fiji	2	1	0	0	∞	∞	∞	∞
Total	300	23	0.380 + 0.036	0.0009 + 0.0009	-31.227	<0.001	0.531	31,600

715

716

717 **Table 2** Gonadotropin-releasing hormone (GnRH) and ribosomal protein S7 molecular
718 diversity indices for *Coris cuvieri* and *Coris gaimard* populations. Location, number of
719 specimens (*n*), number of alleles (N_a), observed heterozygosity (H_O), expected heterozygsity
720 (H_E) and the *P*-value of the exact test of Hardy–Weinberg equilibrium are reported.
721 Maldives (GnRH, *n* = 2; S7, *n* = 0), South Africa (GnRH and S7, *n* = 2), Paracel Islands (GnRH, *n*
722 = 2; S7, *n* = 1), Okinawa (GnRH, *n* = 5; S7, *n* = 1), Fiji (GnRH, *n* = 0; S7, *n* = 1) and the Marshall
723 Islands (GnRH and S7, *n* = 4) are only included in the total values.

Location	GnRH					S7				
C. cuvieri	*n*	N_a	H_O	H_E	*P*	*n*	N_a	H_O	H_E	*P*
Red Sea	9	4	0.375	0.442	0.384	8	7	0.778	0.791	0.742
Djibouti	20	3	0.429	0.408	1.000	19	13	0.944	0.910	0.960
Seychelles	14	5	0.353	0.323	1.000	10	10	0.700	0.884	0.095
Diego Garcia	14	3	0.286	0.442	0.258	12	13	0.917	0.946	0.684
Total	61	7	0.355	0.386	0.760	51	22	0.863	0.903	0.090

C. gaimard

MHI	50	3	0.118	0.180	0.057	58	14	0.741	0.830	0.013
NWHI	34	2	0.152	0.142	1.000	31	13	0.875	0.880	0.519
Johnston	37	2	0.000	0.104	0.001	33	11	0.848	0.846	0.440
Kiribati	26	3	0.308	0.298	0.224	19	14	0.842	0.818	0.721
Moorea	17	3	0.118	0.219	0.177	13	7	0.692	0.720	0.414
Cook Is.	12	2	0.000	0.290	0.006	0	-	-	-	-
Palau	24	2	0.292	0.254	1.000	22	13	0.818	0.851	0.124
Philippines	20	2	0.150	0.224	0.245	0	-	-	-	-
Christmas Is.	38	2	0.262	0.265	1.000	36	21	0.730	0.835	0.001
Cocos Is.	9	2	0.222	0.209	1.000	6	6	0.500	0.818	0.048
Total	278	5	0.158	0.217	<0.001	225	34	0.780	0.850	<0.001

Table 3 Pairwise ϕ_{ST} statistics for *Coris gaimard* cytochrome *c* oxidase subunit I (COI) and ribosomal protein S7 loci. ϕ_{ST} values for COI below diagonal, ϕ_{ST} values for S7 above diagonal. *indicates $P < 0.01$ as per Benjamini and Hochberg (2009) FDR correction for COI and $P < 0.005$ for S7. Bold indicates $P < 0.05$. No S7 sequences were successfully amplified from the Cook Islands or the Philippines. Abbreviations: MHI = Main Hawaiian Islands; NWHI = Northwestern Hawaiian Islands; Johns = Johnston Atoll; Kirib = Kiribati; Moor = Moorea; Cook = Cook Islands; Phil = Philippines; Chris = Christmas Island; Cocos = Cocos-Keeling Islands.

	MHI	NWHI	Johns	Kirib	Moor	Cook	Palau	Phil	Chris	Cocos

MHI	-	-0.005	0.002	**0.033**	**0.056**	-	**0.039**	-	**0.052***	0.044
NWHI	-0.011	-	0.013	0.021	**0.039**	-	**0.026**	-	**0.037**	0.031
Johns	-0.001	-0.004	-	**0.057***	**0.074***	-	**0.056***	-	**0.075***	**0.072**
Kirib	**0.143***	**0.135***	**0.133***	-	-0.019	-	-0.017	-	-0.012	-0.027
Moor	**0.173***	**0.137***	**0.136***	**0.100**	-	-	-0.010	-	0.002	0.268
Cook	0.000	0.000	-0.009	0.077	0.047	-	-	-	-	-
Palau	**0.050***	**0.046***	**0.040***	-0.014	0.037	-0.010	-	-	-0.015	-0.034
Phil	0.037	0.048	0.042	-0.021	0.084	0.040	-0.042	-	-	-
Chris	**0.060***	**0.064***	**0.057***	-0.002	**0.130***	0.022	-0.014	-0.045	-	-0.028
Cocos	0.003	0.002	-0.006	0.057	0.038	0.001	-0.032	0.043	0.012	-

	MHI	NWHI	Johns	Kirib	Moor	Cook	Palau	Phil	Chris	Cocos

Table 4 Pairwise ϕ_{ST} statistics for *Coris gaimard* gonadotropin-releasing hormone (GnRH) locus. ϕ_{ST} values for GnRH below diagonal, *P*-values above diagonal. Bold indicates $P < 0.05$. *P*-values are not significant after correcting for False Discovery Rate (FDR). Abbreviations: MHI = Main Hawaiian Islands; NWHI = Northwestern Hawaiian Islands; Johns = Johnston Atoll; Kirib = Kiribati; Moor = Moorea; Cook = Cook Islands; Phil = Philippines; Chris = Christmas Island; Cocos = Cocos-Keeling Islands.

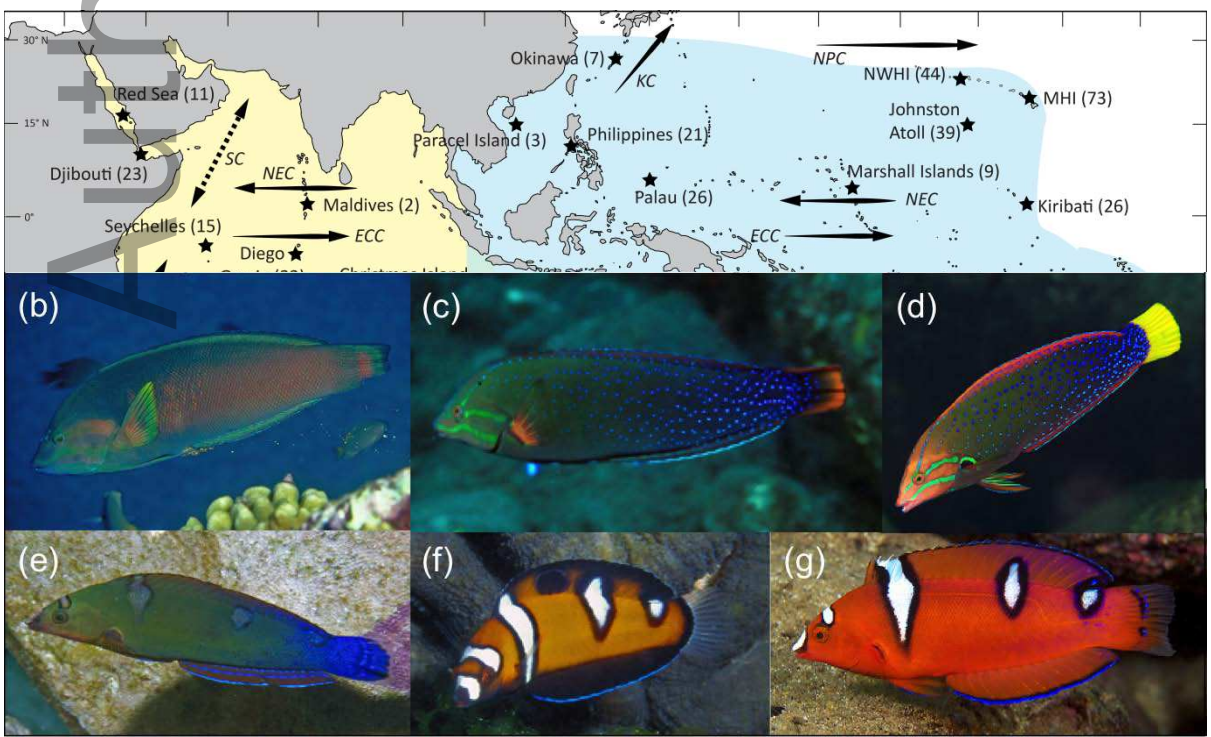
MHI	-	0.780	0.395	0.239	0.478	0.467	0.393	0.763	0.030	1.000
NWHI	-0.010	-	0.732	0.201	0.629	0.236	0.230	0.501	0.032	0.636
Johns	-0.003	-0.011	-	0.059	0.356	0.096	0.109	0.273	0.008	0.600
Kirib	0.006	0.015	0.038	-	0.408	1.000	1.000	0.775	0.399	0.811
Moor	-0.005	-0.006	0.002	0.000	-	0.540	0.396	0.584	0.089	1.000
Cook	0.003	0.020	0.055	-0.030	-0.008	-	1.000	0.719	0.772	0.685
Palau	0.000	0.010	0.033	-0.019	-0.003	-0.031	-	1.000	0.484	1.000
Phil	-0.011	-0.004	0.015	-0.018	-0.011	-0.027	-0.022	-	0.319	1.000
Chris	**0.045**	**0.059**	**0.088**	-0.005	0.030	-0.022	-0.003	0.005	-	0.511
Cocos	-0.031	-0.027	-0.008	-0.030	-0.032	-0.038	-0.034	-0.041	-0.003	-

Figure 1 Map of Indo-Pacific sampling locations for *Coris gaimard* and *C. cuvieri*. (a) Total sample sizes at each location are indicated within the range of each species (yellow shading = *C. cuvieri*; blue shading = *C. gaimard*). (b) *C. cuvieri,* adult male colouration, (c) putative *C. gaimard* x *C. cuvieri* hybrid, adult male colouration; (d) *C. gaimard,* adult male colouration *(*e) *C. cuvieri,* intermediate colouration between juvenile and adult (f) *C. cuvieri* juvenile (g) *C. gaimard* juvenile. Arrows indicate current directions. *SC* = Somali Current (dashed line indicates a changing seasonal current), *NEC* = North Equatorial Current, *ECC* = Equatorial Counter Current, *EACC* = East African Coastal Current, *KC* = Kuroshio Current, *NPC* = North Pacific Current. *Photo credits for b, c, e and f: Robert F. Myers. d, g: Keoki Stender.*

Figure 2 Median joining haplotype network for cytochrome *c* oxidase subunit I (COI) locus. Each circle represents one haplotype, the size of the circle corresponds to the abundance of individuals and the colour indicates the collection location (see legend). Each line and crossbar indicate single mutations. Haplotypes marked as 1 are *Coris cuvieri* haplotypes and

include 4 specimens that had hybrid colouration. Haplotypes marked as 2 are *Coris gaimard* haplotypes and include 6 individuals that had hybrid colouration. Abbreviation: NWHI = Northwestern Hawaiian Islands.

Figure 3 Median joining allele network for *Coris gaimard* and *C. cuvieri* along with suspected hybrids based on the (a) Gonadotropin-releasing hormone (GnRH) locus and (b) Ribosomal protein S7 locus. Each circle represents one allele. The size of the circle is proportional to the abundance of the allele and colour indicates the species (see legend). Each line and crossbar indicate single mutations. Open circles are unsampled haplotypes.

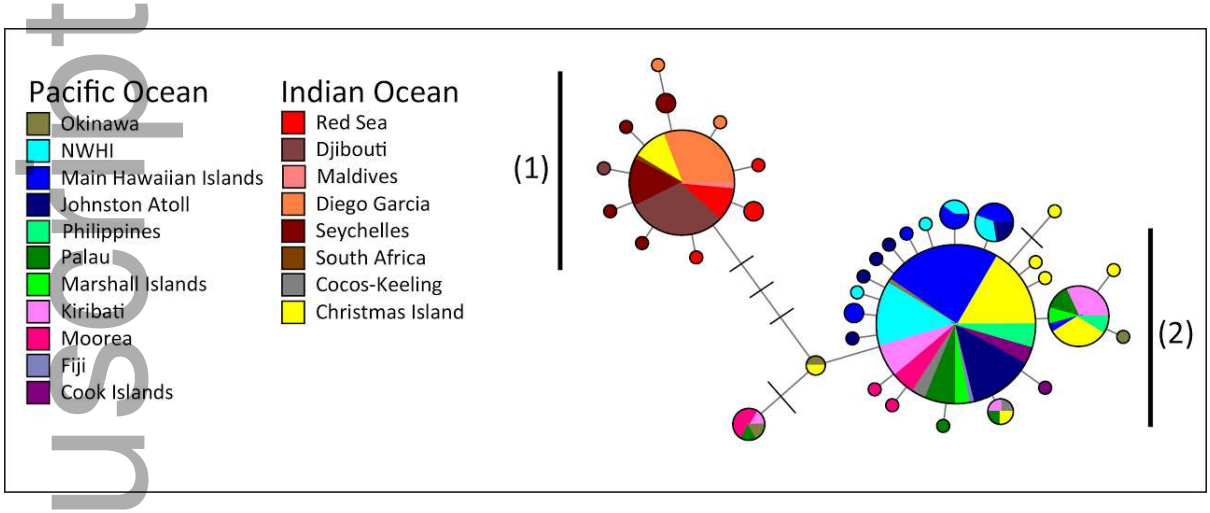


This

792

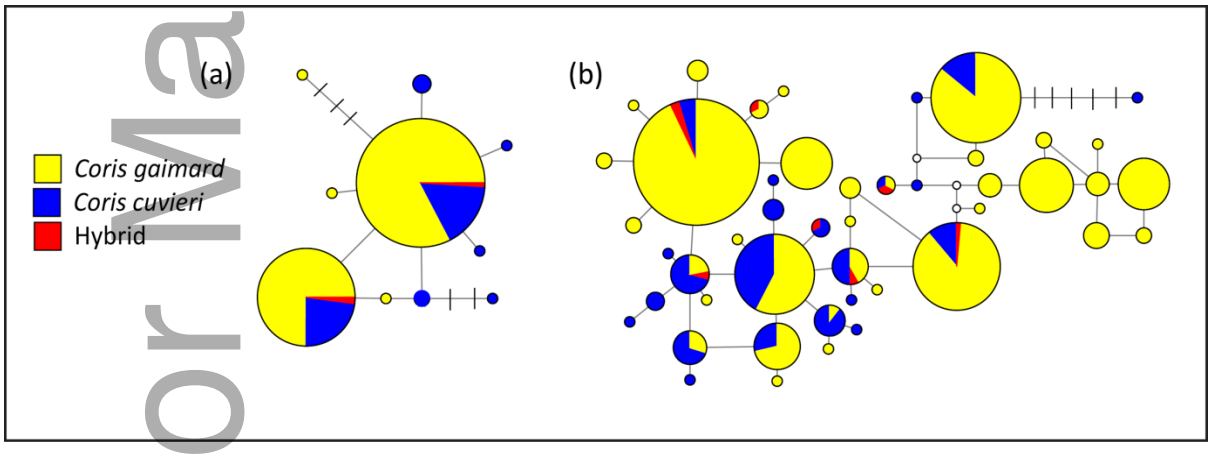
793

794

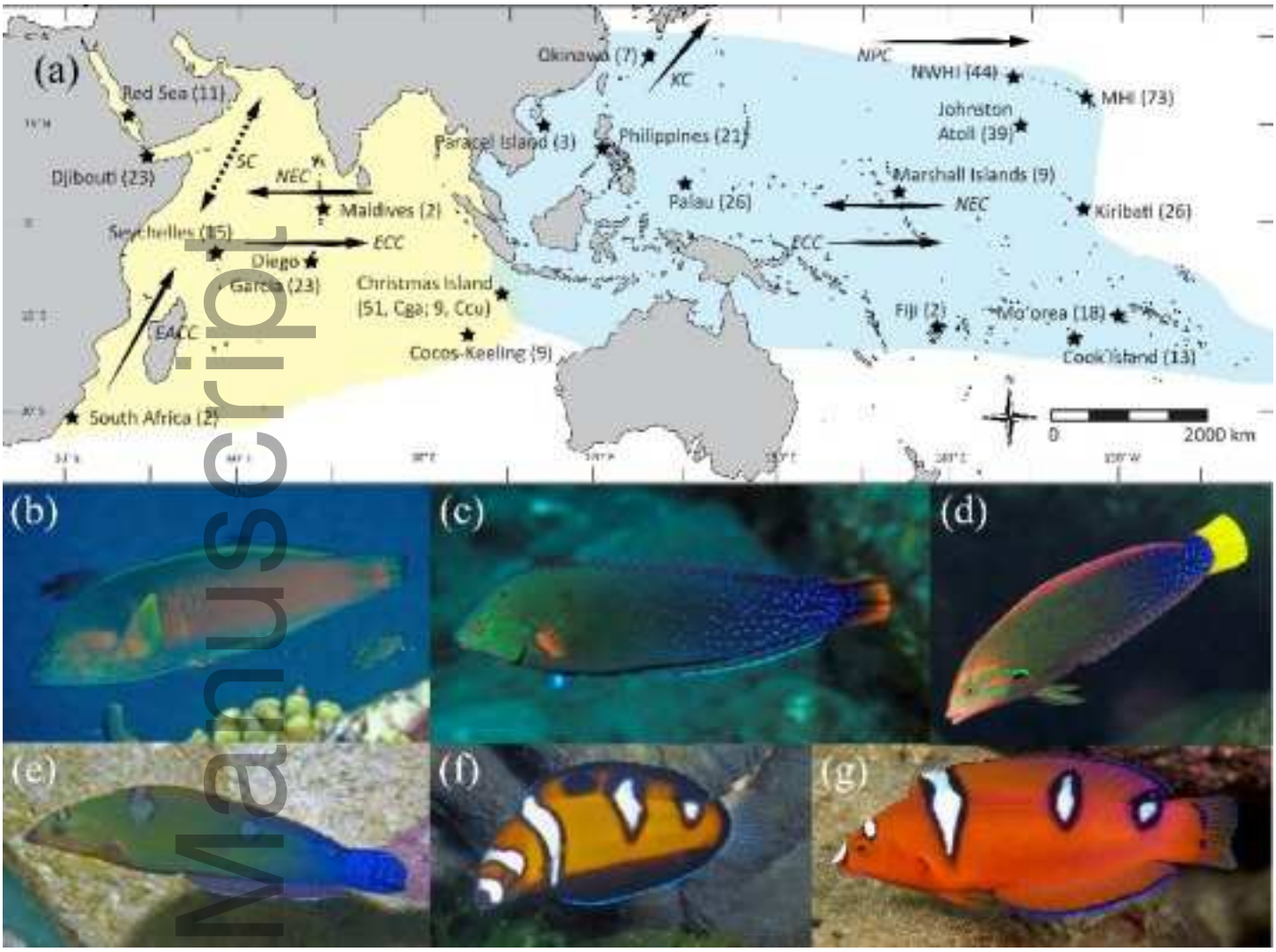


795

796



797



jbi_12712_f1.png