For the virtual issue, "Red Sea and Western Indian Ocean Biogeography"

Original Article

Phylogeography, population structure and evolution of coral-eating butterflyfishes (subgenus *corallochaetodon*)

Ellen Waldrop[1], Jean-Paul A. Hobbs[2], John E. Randall[3], Joseph D. DiBattista[2,4], Luiz A. Rocha[5], Randall K. Kosaki[6], Michael L. Berumen[4] and Brian W. Bowen[1*]

[1]*Hawai'i Institute of Marine Biology, Kane'ohe, HI 96744, USA,* [2]*Department of Environment and Agriculture, Curtin University, PO BOX U1987, Perth, WA 6845, Australia,* [3]*Bishop Museum, Honolulu, HI 96817, USA,* [4]*Red Sea Reearch Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia,* [5]*Section of Ichthyology, California Academy of Sciences, San Francisco, CA 94118, USA,* [6]*Papahānaumokuākea Marine National Monument, NOAA/Daniel K. Inouye Regional Center, Honolulu, HI 96818, USA*

*Correspondence: Brian W. Bowen, Hawai'i Institute of Marine Biology, P.O. Box 1346, Kane'ohe, HI, 96744, USA.
E-mail: bbowen@hawaii.edu

ABSTRACT

Aim This study compares the phylogeography, population structure and evolution of four butterflyfish species in the subgenus *corallochaetodon*, with two widespread species (Indian Ocean - *Chaetodon trifasciatus* and Pacific Ocean - *C. lunulatus*), and two species that are largely restricted to the Red Sea (*C. austriacus*) and northwestern (NW) Indian Ocean (*C. melapterus*). Through extensive geographical coverage of these taxa, we seek to resolve patterns of genetic diversity within and between closely-related butterflyfish species in order to illuminate biogeographical and evolutionary processes.

Location Red Sea, Indian Ocean and Pacific Ocean.

Methods A total of 632 individuals from 24 locations throughout the geographical ranges of all four members of the subgenus *corallochaetodon* were sequenced using a 605 bp fragment (cytochrome *b*) of mtDNA. In addition, 10 microsatellite loci were used to assess population structure in the two widespread species.

Results Phylogenetic reconstruction indicates that the Pacific Ocean *C. lunulatus* diverged from the Indian Ocean *C. trifasciatus* approximately 3 million years ago, while *C. melapterus* and *C. austriacus* comprise a cluster of shared haplotypes derived from *C. trifasciatus* within the last 0.75 Myr. The Pacific *C. lunulatus* had significant population structure at peripheral locations on the eastern edge of its range (French Polynesia, Johnston Atoll, Hawai‘i), and a strong break between two ecoregions of the Hawaiian Archipelago. The Indian Ocean *C. trifasciatus* showed significant structure only at the Chagos Archipelago in the central Indian Ocean, and the two range-restricted species showed no population structure but evidence of recent population expansion.

Main conclusions Patterns of endemism and genetic diversity in *corallochaetodon* butterflyfishes have been shaped by 1) Plio-Pleistocene sea level changes that facilitated evolutionary divergences at biogeographical barriers between Indian and Pacific Oceans, and the Indian Ocean and Red Sea, and 2) semi-permeable oceanographic and ecological barriers working on a shorter timescale. The evolution of range-restricted species (Red Sea and NW Indian Ocean) and isolated populations (Hawai‘i) at peripheral biogeographic provinces indicates that these areas are evolutionary incubators for reef fishes.

Keywords

Biogeography, *Chaetodon austriacus, Chaetodon lunulatus, Chaetodon melapterus, Chaetodon trifasciatus*, microsatellites, mtDNA, reef fish, speciation

INTRODUCTION

How do new species arise in an aquatic medium with high dispersal potential? The Indo-Pacific reef fishes have two biogeographic traits that inform this issue. First, the biodiversity of fishes and other coral-associated species peaks at the central Indo-Australian Archipelago, where Indian and Pacific Ocean faunas overlap (Blum, 1989; Gaither & Rocha, 2013). Second, the highest endemism is in peripheral regions at the ends of the range, including the Red Sea and Hawai‘i (Randall, 1998). Evidence supporting genetic differentiation in peripheral biogeographical regions comes from both locations, which are the western and eastern limits for numerous Indo-Pacific species (DiBattista *et al.*, 2013; Eble *et al.*, 2015). Phylogeographical studies indicate that new species are arising in both the peripheral regions and the biodiversity centre (Bowen *et al.*, 2013). However, few studies have focused on diversification in the Red Sea and northwestern (NW) Indian Ocean.

The well-resolved phylogeny of butterflyfishes (family Chaetodontidae), has made this group an appropriate model for understanding the evolution of reef fishes (Fessler & Westneat, 2007; Cowman & Bellwood, 2013; Hodge *et al*., 2014). Butterflyfishes embody the primary biogeographic patterns outlined above, with greatest diversity in the Indo-Australian Archipelago and highest endemism in peripheral areas. The Red Sea and adjacent Gulf of Aden has 32% endemism in butterflyfishes, compared to 13% in Hawai‘i and < 10% elsewhere in the Indo-Pacific (Randall, 2007; DiBattista *et al.*, in review). Understanding how the highest levels of endemism arose far from the center of diversity remains an enigma. Biogeographical barriers at these locations may have created isolated populations or endemic species depending on the divergence time (Briggs & Bowen, 2013).

Among butterflyfishes, the subgenus *corallochaetodon* contains four corallivorous species that have mostly allopatric distributions with narrow areas of overlap on the range edges (Fig. 1). *Chaetodon*

lunulatus Quoy & Gaimard, 1824 occurs throughout the Pacific Ocean from Hawai‘i and the Tuamotu Islands westward to Indonesia and the eastern Indian Ocean (Christmas Island), while *Chaetodon trifasciatus* Park, 1797 is distributed in the Indian Ocean from Indonesia and Christmas Island to East Africa, but is not known from the Red Sea (Allen *et al*., 1998). *C. lunulatus* and *C. trifasciatus* may be Indian-Pacific Ocean sister species that diverged during Plio-Pleistocene sea level changes that created the transient Sunda Shelf Barrier (Hsu *et al*., 2007). *Chaetodon melapterus* Guichenot, 1863 is restricted to the Arabian Gulf, Gulf of Oman, Gulf of Aden and the southern Red Sea, while *Chaetodon austriacus* Rüppell, 1836 occurs predominantly in the northern and central Red Sea (Zekeria *et al.*, 2005), with rare records in the southern Red Sea and adjacent Arabian Sea (DiBattista *et al.*, in review). It is unknown if the two range-restricted species (*C. melapterus* and *C. austriacus*) arose independently, and whether they evolved from the widespread Indian Ocean species *C. trifasciatus*, as geography would indicate. Thus the subgenus *corallochaetodon* provides the opportunity to determine how the speciation of butterflyfishes in peripheral locations (*C. melapterus* and *C. austriacus*) compares to that in the center of diversity (*C. lunulatus* and *C. trifasciatus*).

This study is motivated by four primary questions. First, what is the evolutionary history of the subgenus *corallochaetodon*? Second, what are the geographical patterns of genetic diversity within and between species? Third, what is the population structure (as revealed by mtDNA) of all four species across their geographical ranges? Fourth, what is the fine-scale population structure (as revealed by microsatellite DNA) in the two widespread species (*C. lunulatus* and *C. trifaciatus*), and is there evidence of peripheral speciation? These genetic patterns can illuminate the origins of marine biodiversity, and the measures that would conserve building blocks of future biodiversity.

MATERIALS AND METHODS

Sample Collection

Tissue specimens (fin clips or gill filament) were obtained using polespears whilst SCUBA diving at 24 locations across the Indo-Pacific (including the Red Sea) from 2005 to 2013 (*C. lunulatus N* = 603, *C. trifasciatus N* = 143, *C. melapterus N* = 95, *C. austriacus N* = 30) (Table 1). *Chaetodon lunulatus* was intensively sampled in the Hawaiian Archipelago to assess connectivity across this 2600 km island chain. All tissues were preserved in a saturated salt DMSO solution (Seutin *et al*., 1991). DNA was extracted using a "HotSHOT" protocol (Meeker *et al*., 2007), and aliquots were stored at -20 ˚C.

Mitochondrial DNA Sequencing

A 605 base pair (bp) segment of mtDNA cytochrome *b* (cyt *b*) gene was resolved for all specimens. Details of the PCR methodology are available in Box S1 in Appendix S1 and Waldrop (2014). The cyt *b* data comprises a single locus but offers the advantage of haploid inheritance, lack of recombination, comparison to existing studies and availability of universal primers for efficient production of sequence data. Unique mtDNA cyt *b* haplotypes are deposited in GenBank under accession numbers KP241594 to KP241672.

Phylogenetic relationships

Phylogenetic relationships were examined among the four species by constructing neighbour-joining (NJ), maximum-likelihood (ML) and maximum-parsimony (MP) trees from the cyt *b* haplotypes of all individuals (PAUP* implemented in Geneious Pro 6.0.6 and MEGA 5.2.2; Swofford, 2003; Drummond *et al.*, 2010; Tamura *et al*., 2011). Bootstrap support values were calculated using default settings with 10,000 replicates. A single *Chaetodon vagabundus* Linnaeus, 1758 sample (Genbank accession numbers: JF458006) was used to root trees. For simplicity, a subset of unique haplotypes was used to create the final tree. An unrooted network of haplotypes was also assembled using a median-

joining algorithm and default settings in NETWORK 4.5.1.0 (Bandelt *et al*., 1999). Molecular clock rate is provisionally estimated at 2% per Myr (between lineages) for the cyt *b* gene (Bowen *et al*., 2001; Reece *et al*., 2011). Evolutionary distances among lineages were calculated with the Tamura-Nei model and 1,000 bootstrap replicates in MEGA.

Population structure for mtDNA

An Akaike Information Criterion (AIC) test in jModelTest 2.1.3 (Posada, 2008) was used to determine the best nucleotide substitution model for each species. The HKY model (Hasegawa *et al*., 1985) was selected for *C. lunulatus*, *C. trifasciatus* and *C. austriacus*, and TrN+G (Tamura & Nei, 1993) was selected for *C. melapterus*. The TrN+G is the only one of these models available in analytical software and was selected for all phylogeographical inferences. Arlequin 3.5.1.3 (Excoffier *et al*., 2005) was used to calculate haplotype (*h)* and nucleotide diversity *(π)*, Fu's *Fs* test of neutrality (Fu, 1997) and apply an analysis of molecular variance (AMOVA; Excoffier *et al*., 1992) to test for patterns of population structure; tests were run for each species separately. Samples with $N < 5$ were excluded from all population-level analyses and pooled into their respective larger sampling locations to provide adequate statistical power. Hawaiian specimens of *C. lunulatus* were subdivided into the Main Hawaiian Islands (MHI, high islands) and Northwestern Hawaiian Islands (NWHI, low islands and atolls) to test for genetic structure within the archipelago. *C. trifasciatus* specimens from the eastern Indian Ocean (Cocos-Keeling Islands and adjacent Christmas Island) were pooled to increase statistical power as they were indistinguishable in preliminary analyses.

Population structure - microsatellites

153 Microsatellite primers were designed for *C. lunulatus* by Lawton *et al.* (2010; 2011). Here the
154 widespread *C. lunulatus* and *C. trifasciatus* were genotyped at 10 loci (Table S1 in Appendix S1). The
155 range-restricted *C. melapterus* and *C. austriacus* were not genotyped because large samples were not
156 available, finances were limited and cross-species applications can be complicated by allele dropout,
157 homoplasy and other problems (see Selkoe & Toonen, 2006). Details of PCR amplifications are
158 available in Appendix S1 and Waldrop (2014). Initially specimens from Hawai‘i were separated into
159 individual sampling locations by island. However mtDNA data revealed a genetic break between the
160 MHI and NWHI concordant with a multi-species connectivity study (Toonen *et al.*, 2011). For
161 subsequent analyses, Hawai‘i was partitioned into two groups; MHI and NWHI. However, a full
162 comparison among Hawaiian sample sites is provided in Table S1 in Appendix S2.
163

164 For each locus the mean number of alleles (N_A), observed (H_O) and expected (H_E) heterozygosities,
165 departure from Hardy-Weinberg proportions (HWE) and linkage disequilibrium (LD) were assessed
166 with GENEPOP 4.2 (Raymond & Rousset, 1995). MICRO-CHECKER 2.2.3 was used to identify null
167 alleles and excessive stutter peaks (van Oosterhout *et al.*, 2004), and significance levels for multiple
168 comparisons were adjusted using the sequential Bonferonni correction. GENODIVE 2.0b23 (Meirmans
169 & Tienderen, 2004) was used to estimate population structure for each species. STRUCTURE 2.3.4
170 was used to assign individuals to distinct genetic clusters (populations) without presumption of
171 predefined geographical locations (Pritchard *et al.*, 2000). The most likely number of clusters was
172 identified based on the probability of $K = 1$ to $K = 12$ or $K = 1$ to $K = 4$ for *C. lunulatus* and *C.*
173 *trifasciatus*, respectively. Analyses were repeated five times and averaged. Each replicate run consisted
174 of 1,000,000 MCMC repetitions, a burn-in of 10,000 iterations and assumed correlated allele
175 frequencies with admixed populations (as per DiBattista *et al.*, 2012). STRUCTURE HARVESTER
176 0.6.93 was used to determine most likely value of *K* following Evanno *et al.* (2005) to visualize

likelihood values and the number of groups that best fit the data (Earl & von Holdt, 2012).

RESULTS

Phylogenetic relationships

The authors recognize the limitations of a single-locus phylogeny, and so here we provide the mtDNA results as an initial hypothesis of relationships among the four species. All tree-building methods used to analyze the mtDNA cyt *b* fragment (605 bp) produced nearly identical tree topologies with bootstrap support values for species level relationships of 80 to 100% (Fig. 2). The primary feature of this phylogeny is a bifurcation with $d = 0.06$ sequence divergence between Pacific Ocean *C. lunulatus* and the Indian Ocean *C. trifasciatus*. The two range-restricted species, *C. melapterus* and *C. austriacus*, are more closely related to the Indian Ocean species ($d = 0.015$). However, they did not form monophyletic groups, and share the most common haplotype (Fig. 2). The relationship within the subgenus *corallochaetodon* is apparent in the parsimony network (Fig. 3), where Pacific Ocean *C. lunulatus* and Indian Ocean *C. trifasciatus* are separated by 28 diagnostic nucleotide substitutions, and the *C. melapterus–C. austriacus* cluster is separated from *C. trifasciatus* by three diagnostic nucleotide substitutions.

Genetic diversity

Haplotype diversity within each species was moderate to high (*C. lunulatus h* = 0.45 to 0.87; *C. trifasciatus h* = 0.67 to 0.80; *C. melapterus h* = 0.63 to 0.78; *C. austriacus h* = 0.84 to 0.87; Table 1). For the species with the largest geographic range (*C. lunulatus*), haplotype diversity was highest at the peripheral location on the western edge of its range (Christmas Island), and was generally lowest at peripheral locations on the eastern edge of its range (Johnston Atoll, Main Hawaiian Islands - MHI, Northwestern Hawaiian Islands - NWHI). For *C. trifasciatus*, haplotype diversities are similar across

the range. In the two range-restricted species (*C. melapterus*, and *C. austriacus*), haplotype diversity was lower at one sampled location (Table 1). Nucleotide diversity was low for all species (*C. lunulatus* π = 0.001 to 0.005; *C. trifasciatus* π = 0.001 to 0.088; *C. melapterus* π = 0.000 to 0.001; *C. austriacus* π = 0.000 to 0.002; Table 1), indicating a cluster of closely-related haplotypes within each species.

For the two widespread species, only one of the 17 sample locations was significant for Fu's *Fs* (*C. trifasciatus* at Diego Garcia). For the two range-restricted species, tests for Fu's *Fs* could only be conducted on samples from five locations and all produced significant negative values: *C. melapterus* at Maskali, Obock and Oman; *C. austriacus* at Jazirat Baraqan and Yanbu (Table 1).

Population structure (mtDNA)

Significant population structure was observed in *C. lunulatus* (overall Φ_{ST} = 0.27; $P < 0.001$). In comparisons among sample locations, 30 out of 78 pairwise comparisons were statistically significant ($P < 0.05$; Table 2). Five locations accounted for all the significant comparisons: Fiji with 6 out of 12 significant comparisons, Johnston Atoll with 3 out of 12 significant comparisons, Mo'orea (French Polynesia) with 12 out of 12 significant comparisons, MHI with 5 out of 12 significant comparisons and the NWHI with 12 out of 12 significant comparisons (Table 2). Within the Hawaiian Archipelago, there were 13 out of 28 significant comparisons among sample locations (Table S1 in Appendix S2). All of the significant comparisons were among the three southernmost sampled locations (Hawai'i Island, O'ahu and French Frigate Shoals) and the most northern sample location (Kure Atoll).

No significant structure overall or significant pairwise comparisons were detected among four locations in *C. trifasciatus* (Φ_{ST} = 0.01; P = 0.50), four locations in *C. melapterus* (Φ_{ST} = 0.01; P = 0.16), or three locations in *C. austriacus* (Φ_{ST} = 0.04; P = 0.21) (Table 3). However, *C. melapterus* and *C. austriacus*

225 were significantly isolated at a population level (Φ_{ST} = 0.06; P = 0.001). Notably, we did not sample *C.*
226 *melapterus* in the Arabian Gulf and along the Somalian coastline due to logistical limitations; dditional
227 sampling in these regions could change conclusions about population structure.
228
229 **Population structure (msatDNA) within *C. lunulatus* and *C. trifasciatus***
230 Significant population structure was also detected for *C. lunulatus* using msatDNA (F_{ST} = 0.05, P =
231 0.001). The msatDNA results were similar to that of mtDNA with most of the significant pairwise
232 comparisons involving locations on the eastern edge of the geographic range: Johnston Atoll, Mo‘orea,
233 MHI and the NWHI. Microsatellite allele frequencies were significantly different in 49 out of 91
234 comparisons for *C. lunulatus* (Table 4; see also Table S1 in Appendix S2).
235
236 For *C. lunulatus*, STRUCTURE identified mean probabilities as being highest at K = 3 (Fig. 4), which
237 was verified using STRUCTURE HARVESTER (Fig. S1 in Appendix S2). One widespread population
238 spanned locations from the western range edge (Christmas Island and Indonesia) eastward to Kiribati in
239 the central Pacific Ocean. The second population was comprised predominately of individuals from
240 isolated locations on the eastern range edge: Johnston Atoll, MHI and the NWHI. The third population
241 was largely restricted to the NWHI.
242
243 The msatDNA data revealed low but significant population structure for *C. trifasciatus* (F_{ST} = 0.003, P
244 = 0.03). Microsatellite allele frequencies were significantly different in three out of six comparisons
245 (Table 5), between Diego Garcia and all the other sampled locations (Seychelles, Christmas Island and
246 Indonesia). Microsatellite statistics for each location and both species are provided in Table S2 in
247 Appendix S2. STRUCTURE identified mean probabilities as being highest at K = 2 (Fig. 5), which was
248 consistent with the results from STRUCTURE HARVESTER (Fig. S2 in Appendix S2), indicating

249 isolation of Diego Garcia but no distinction of samples from the east (Christmas Island, Indonesia) and
250 west (Seychelles) of this remote location in the Chagos Archipelago. Overall, there was no consistent
251 evidence for departure from HWE, linkage disequilibrium or null alleles across all sampled locations in
252 both species.
253
254 **DISCUSSION**
255 **Phylogenetic relationships**
256 The primary phylogenetic feature of the subgenus *corallochaetodon* is mtDNA sequence divergence of
257 d = 0.06 between Indian Ocean *C. trifasciatus* and Pacific *C. lunulatus*. Based on the conventional
258 molecular clock of 2% per Myr, this corresponds to approximately three Myr of separation (Table S3 in
259 Appendix S2) (consistent with Hsu *et al*., 2007; Bellwood *et al*., 2010), which is close to the onset of
260 modern glacial cycles at 2.6 to 2.8 Ma (Dwyer *et al*., 1995; Williams *et al*., 1997). The shallow Sunda
261 Shelf is exposed during glacial periods with low sea levels, forming land bridges through the
262 Indonesian Archipelago that restricted exchange between the Indian and Pacific Oceans (Randall,
263 1998; Rocha *et al*., 2007). This indicates that transient allopatry had a role in the formation of this
264 species pair, a process that is apparent (or suspected) in other Indian-Pacific species pairs (Gaither &
265 Rocha, 2013).
266
267 A divergence time of approximately three Myr for *C. trifasciatus* and *C. lunulatus* falls within the
268 range of divergence times (0.3 – 6.6 Myr) for other Indian and Pacific sister species of reef fishes
269 (Gaither & Rocha, 2013). However, divergence times in other Indian and Pacific Ocean butterflyfish
270 sister species tends to be less (0.3 – 1.4 Myr) (Fessler & Westneat 2007; Hsu *et al.*, 2007; Bellwood *et*
271 *al*., 2010; DiBattista *et al*., 2012). Variation in divergence times may be due to a number of factors
272 including: 1) potential differences in mutation rates; 2) the intermittency of the Sunda Shelf Barrier

during the Pleistocene due to repeated glacial cycles (i.e. different species pairs diverged at different low sea level stands); and 3) the conditions determining secondary contact and reproductive isolation affected species differently.

The range-restricted *C. austriacus* and *C. melapterus* share a common haplotype, and are closely affiliated with *C. trifasciatus* (d = 0.015). The divergence between *C. trifasciatus* and the range-restricted species is approximately 0.75 Myr (Table S3 in Appendix S2), which corresponds with Pleistocene sea level changes that repeatedly isolated the Red Sea region from the Indian Ocean (Fig. 1; Blum, 1989; DiBattista *et al.*, 2013). Furthermore, strong upwelling in the NW Indian Ocean (off the southern Oman coast) may facilitate allopatric divergence between species from the Indian Ocean (e.g. *C. trifasciatus*) and Red Sea to Arabian Gulf region (*C. austriacus* and *C. melapterus*).

While *C. austriacus* and *C. melapterus* are not monophyletic, these two putative species are genetically distinct at a population level (Φ_{ST} = 0.06; P = 0.001) indicating either early stages of speciation or distinct colour morphs separated by habitat discontinuities. This finding should be interpreted in light of the relatively recent origins of reef faunas inhabiting the Red Sea (DiBattista *et al.*, 2013) and Arabian Gulf (Sheppard *et al*., 2010). Estimated time since divergence is approximately 50 kyr, and was likely initiated by vicariant isolation at the Strait of Bab al Mandab (at the mouth of the Red Sea – Fig. 1). This barrier flooded about 20 ka, and *C. austriacus* and *C. melapterus* now have limited contact in the southern Red Sea (Randall, 1994), a region characterised by changes in environmental conditions (e.g. salinity, temperature, nutrients: Kemp, 1998; Sheppard, 1998) that are reflected in the fish community (Roberts *et al*., 1992; DiBattista *et al*., in review). Given that *C. austriacus* and *C. melapterus* inhabit different environmental conditions on either side of this area, successful colonisation across this potential barrier may be limited, thereby facilitating divergence. When the two

297 species come into contact, differences in colouration and assortative mating may maintain reproductive
298 isolation (McMillan *et al*., 1999).
299
300 The distribution of all four sister species overlap at their range edges, at (or adjacent to)
301 biogeographical barriers (Fig. 1). In the eastern Indian Ocean, cohabitation and a breakdown in
302 assortative mating between *C. lunulatus* and *C. trifasciatus* at Christmas Island has led to hybridisation
303 (Hobbs *et al*., 2009; Montanari *et al.,* 2014); however, there has only been limited and localised
304 introgression between the species. In the western Indian Ocean, *C. trifasciatus* and *C. melapterus*
305 hybridise at Socotra, with some evidence of introgression beyond this hybrid zone in Djibouti
306 (DiBattista *et al*., 2015). In the southern Red Sea, *C. austriacus* and *C. melapterus* cohabit and
307 potentially hybridise (Randall, 1994; Kuiter, 2002), but the former is considered rare in this
308 understudied region (Righton e*t al.,* 1996). This pattern of decreasing hybridisation and introgression
309 with increasing divergence time is consistent with other butterflyfish studies (Montanari *et al*., 2014).
310 Overall, it appears that Plio-Pleistocene sea level changes have facilitated allopatric speciation in both
311 the butterflyfish centers of diversity (Indonesia) and peripheral areas (Red Sea). Secondary contact and
312 hybridisation could erode species boundaries (Coleman *et al.,* 2014); however, abrupt differences in
313 environmental conditions across areas of secondary contact could facilitate evolutionary divergence.
314
315 **Genetic diversity**
316 Although the geographical ranges of the four species in the subgenus *corallochaetodon* vary by an
317 order of magnitude, there was no obvious relationship between haplotype diversity and range size.
318 Terrestrial studies commonly find low haplotype diversity in range-restricted endemics (Frankham,
319 1998). However, endemic reef fishes can have population sizes numbering in the millions (Hobbs *et*
320 *al*., 2011) and this may explain why they have haplotype diversities similar to widespread species (Eble

321 *et al.*, 2009; Hobbs *et al.*, 2013; Delrieu-Trottin *et al.*, 2014). Excluding the Arabian Gulf, where
322 atypical conditions have resulted in an unusually low abundance and diversity of butterflyfishes
323 (Pratchett *et al*., 2013), *C. austriacus* and *C. melapterus* are the most common butterflyfish species in
324 their respective ranges (Berumen & Hobbs, unpub. data). Therefore, the large population sizes of the
325 range-restricted *C. austriacus* and *C. melapterus* would help generate and maintain high haplotype
326 diversity. Nearly all the populations of the two restricted-range species had significant negative Fu's *Fs*
327 values. Therefore, it appears that *C. austriacus* and *C. melapterus* have undergone recent population
328 expansion.
329
330 **Population structure - mtDNA**
331 Contrary to a hypothesis proposed by Eble *et al*. (2009), range size does not always predict genetic
332 structure. Data from the wide-ranging *C. lunulatus* (Pacific) indicates strong population structure,
333 whereas the sister species *C. trifasciatus* (Indian) showed significant genetic structure only at Diego
334 Garcia (Chagos Archipelago). Data from the two range-restricted species, *C. austriacus* and *C.*
335 *melapterus*, detected no population structure based on our approach, which may indicate that each
336 represents a single panmictic population. This can be explained by their limited distributions in the NW
337 Indian Ocean, with no apparent biogeographical barriers within each range.
338
339 *Corallochaetodon* mtDNA sequence data revealed that range size was not related to genetic population
340 structure, which is a proxy for realised dispersal ability (Eble *et al.*, 2009). The widespread *C. lunulatus*
341 showed significant population structure at eastern peripheral locations, consistent with known
342 distributional barriers (Blum, 1989; Hsu *et al.*, 2007). The distinction of the Mo'orea population of *C.*
343 *lunulatus* (Lawton *et al.*, 2011; this study) is concordant with other Pacific Ocean species and may be
344 caused by isolating oceanographic currents (Gaither *et al.*, 2010; Eble *et al*., 2011). The isolation of

345 Johnston Atoll indicates that the pelagic larval duration (~35 days: Soeparno *et al*., 2012) of *C.*
346 *lunulatus* is insufficient to make the 40 to 50 day transit to the nearest reef (Hawaiian Archipelago)
347 (Kobayashi, 2006).
348
349 Population differentiation between Hawai‘i and other Pacific locations has been reported in many other
350 reef fishes (Leray *et al*., 2010; DiBattista *et al*., 2011; Gaither *et al*., 2011a; Szabo *et al*., 2014;
351 Fernandez-Silva *et al*., in press; Ahti *et al*., in review). The recurrent trend of genetic distinctness in
352 this region can be attributed to three factors: (1) isolation due to location and oceanographic currents,
353 (2) dispersal characteristics of the fishes and (3) adaptation to environmental conditions in Hawai‘i
354 (Hourigan & Reese, 1987). Widespread reef fishes usually exhibit genetic homogeneity within the
355 Hawaiian archipelago (Craig *et al*., 2007; Eble *et al*., 2009; Gaither *et al*., 2010, 2011a, b; Reece *et al*.,
356 2011; DiBattista *et al*., 2011, 2012; Ludt *et al*., 2012); however, the genetic differentiation of *C.*
357 *lunulatus* across the archipelago (between the low islands of the NWHI and the high volcanic islands of
358 the MHI) is more typical of endemic reef fishes and invertebrates (Eble *et al*., 2009; Craig *et al*., 2010;
359 Toonen *et al*. 2011).
360
361 **Population structure – msatDNA**
362 Investigation of fine-scale population structure in the two widespread species using msatDNA revealed
363 patterns similar to the mtDNA with *C. trifasciatus* exhibiting low structure, whereas *C. lunulatus* had
364 more pronounced structure. For *C. trifasciatus*, the msatDNA differed from mtDNA results in one
365 point –the former support the genetic isolation of Diego Garcia (Chagos Archipelago) in the central
366 Indian Ocean. The population genetic separation of Chagos has been observed in other reef fauna
367 (Gaither *et al*., 2010; Eble *et al*., 2011; Vogler *et al*., 2012) and may be related to seasonal monsoon-

368 driven currents that switch direction between easterly and westerly, possibly limiting larval dispersal to
369 this location (Sheppard *et al*., 2012).
370
371 MsatDNA analyses for *C. lunulatus* were consistent with the mtDNA results in indicating divergent
372 populations at peripheral locations on the eastern range edge: Mo‘orea, Johnston Atoll, MHI and
373 NWHI. The majority of the geographic range of *C. lunulatus* is comprised of relatively close islands
374 and reefs throughout the Central-West Pacific; however, the large distance and prevailing currents
375 work against colonisation of Hawai‘i and French Polynesia, thus explaining the genetic distinctness of
376 populations at these peripheral locations (Hourigan & Reese, 1987; Gaither *et al*., 2010). This isolation
377 is the starting point for peripheral speciation, explaining why Hawai‘i has one of the highest levels of
378 reef fish endemism in the world (Randall, 2007).
379
380 An interesting outcome for *C. lunulatus* is the population separation between the high islands of the
381 MHI and the low islands and atolls of the NWHI; *C. lunulatus* is the first widespread reef fish to show
382 strong population structure across the Hawaiian Archipelago. Part of the explanation may be habitat
383 preference: this species uses sheltered, coral-rich areas and the lack of this habitat between MHI and
384 NWHI may explain the genetic break. Indeed, at the MHI region adjacent to this break (Kaua‘i),
385 previous transect data (unpub. data) and our own efforts indicate a near absence of *C. lunulatus*.
386 Another part of the explanation may include Johnston Atoll to the south. Johnston has long been
387 postulated to be a gateway into Hawai‘i (Hourigan & Reece, 1987), and STRUCTURE analysis shows
388 an affiliation between Johnston and the MHI, to the exclusion of the NWHI (Fig. 4). This invokes the
389 possibility that Hawai‘i was colonized twice, possibly from different sources.
390
391 **Conclusion**

We conclude that Plio-Pleistocene sea level changes have influenced speciation at both the center of diversity and peripheral areas for butterflyfishes of the subgenus *corallochaetodon*. Evolutionary divergence among *corallochaetodon* species may have been initiated along the intermittent biogeographical barriers between Indian and Pacific Oceans, and between the Indian Ocean and Red Sea. Phylogenetic analyses revealed that the two species restricted to the Red Sea to Arabian Sea region are indistinguishable at cyt *b*. There was no evidence that the size of the geographic range is related to genetic diversity or population structure. Genetic diversity decreases from west to east for the widespread *C. lunulatus*, but there are no patterns for the other three species. The two range-restricted species appear to have undergone recent population expansion and exhibit no population structure, while the widespread Indian Ocean species (*C. trifasciatus*) showed little population structure, which is likely attributed to variable local conditions (e.g. seasonal monsoon currents). Peripheral populations on the eastern range edge of the widespread Pacific species *C. lunulatus* were genetically distinct from populations in the center of the range. The recent evolution of *C. melapterus* and *C. austriacus* in the Red Sea to Arabian Sea region, and genetic distinctness of peripheral populations of the widespread *C. lunulatus*, indicate that such peripheral marine habitats can be engines of biodiversity (Bowen *et al*., 2013). Thus peripheral speciation (through isolation and vicariant events) would help explain why the Red Sea and Hawai‘i, at opposite extremes of the Indo-Pacific ranges, are endemic hotspots for reef fishes.

ACKNOWLEDGMENTS

For assistance with field work and collections, we thank Alexander Alfonso, Senifa Annadale, Kim Anderson, Paul H. Barber, W.K. Chan, Howard Choat, Richard Coleman, Pat and Laura Colin, Greg Concepcion, Joshua Copus, Matthew Craig, Toby Daly-Engel, Nancy Daschbach, Joshua A. Drew, John L. Earle, Jeff Eble, Kevin Flanagan, Michelle Gaither, Brian Greene, Matthew Iacchei, Stephen

Karl, Jonathan Mee, Carl Meyer, Darren Okimoto, Yannis Papastamatiou, David Pence, Mark Priest, Jon Puritz, Richard Pyle, Joshua Reece, D. Ross Robertson, Nick Russo, Jennifer Schultz, Charles Sheppard, Derek Skillings, Derek Smith, Zoltan Szabo, Sue Taei, Kim Tenggardjaja, Tukabu Teroroko, Robert Thorn, Allen Tom, Bill Walsh, Christie Wilcox, Ivor Williams, Jill Zamzow and the officers and crew of the NOAA ship Hi‘ialakai. For logistic support and advice, we thank Robert Toonen, Robert Thomson, Stephen Karl, Jo-Ann Leong, David Pence, Eric Mason at Dream Divers, Nicolas Prévot at Dolphin Divers and the crew of the M/V Deli in Djibouti, the Ministry of Agriculture and Fisheries in Oman including Abdul Karim, Jason Jones, Narineh Nazarian, Iria Fernandez-Silva, Zac Forsman, Jennier McIlwain and Amy Eggers. For institutional support, we thank the Papahānaumokuākea Marine National Monument, University of Hawai‘i Diving Safety Program, Hawai‘i Department of Land and Natural Resources, the KAUST Coastal and Marine Resources Core Lab and Amr Gusti, Conservation International, Coral Reef Research Foundation, Phoenix Island Protected Area, Dept. of the Environment (Australia), Government of Kiribati, Administration of the British Indian Ocean Territories, Dept. of Marine and Wildlife Resources (American Samoa), Government of Fiji and the Chiefs and people of Wagamimi, Tavewa, and Yasawas villages, Government of French Polynesia and U.S. Fish and Wildlife Service (Johnston Atoll). We thank Zoltan Szabo, the Center for Genomics, Proteomics, and Bioinformatics at the University of Hawai’i at Manoa and the KAUST Bioscience Core Facility for their assistance with DNA sequencing. Thanks to Keoki Stender for sharing photographs. We thank editor Michelle Gaither and two anonymous reviewers for valuable comments that improved the manuscript. This project was funded by the National Science Foundation Grant OCE-0929031 to B.W.B., Elizabeth Alison Kay Endowed Award, University of Hawai‘i Graduate Student Organization, NOAA National Marine Sanctuaries Program MOA No. 2005-008/66882 to R.J. Toonen, National Geographic Society Grant 9024-11 to J.D.D. and KAUST

Office of Competitive Research Funds (OCRF) under Award No. CRG-1-2012-BER-002 and baseline research funds to M.L.B.

LITERATURE CITED

Ahti, A.P., Coleman, R.R., DiBattista, J.D., Berumen, M.L., Rocha, L.A. & Bowen B.W. (in review) Phylogeography of Indo_Pacific reef fishes: Sister wrasses *Coris gaimard* and *C. cuvieri* in the Red Sea, Indian Ocean, and Pacific Ocean. *Journal of Biogeography.*

Allen, G.R., Steene, R.C. & Allen, M. (1998) *A Guide to Angelfishes & Butterflyfishes*. Odyssey Publishing/Tropical Reef Research, Perth.

Bandelt, H.J., Forster, P., & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bellwood, D.R., Klanten, S., Cowman, P.F., Pratchett, M.S., Konow, N. & van Herwerden, L. (2010) Evolutionary history of the butterflyfishes (f: Chaetodontidae) and the rise of coral feeding fishes. *Journal of Evolutionary Biology*, **23**, 335-349.

Blum, S.D. (1989) Biogeography of the Chaetodontidae: an analysis of allopatry among closely related species. *Environmental Biology of Fishes*, **25**, 9-31.

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S. & Robertson, D.R. (2001) Phylogeography of the trumpetfishes (*Aulostomus*): Ring species complex on a global scale. *Evolution*, **55**, 1029-1039.

463 Bowen, B.W., Rocha, L.A., Toonen, R.J., Karl, S.A. & The ToBo Laboratory. (2013) The origins of
464 tropical marine biodiversity. *Trends in Ecology & Evolution*, **28**, 359-366.
465
466 Briggs, J.C. & Bowen, B.W. (2013) Evolutionary patterns: Marine shelf habitat. *Journal of*
467 *Biogeography*, **40**, 1023–1035. doi:10.1111/jbi.12082.
468
469 Coleman, R.R., Gaither, M.R., Kimokeo, B., Stanton, F.G., Bowen, B.W. & Toonen, R.J. (2014)
470 Large-scale introduction of the Indo-Pacific damselfish *Abudefduf vaigiensis* into Hawai'i promotes
471 genetic swamping of the endemic congener *A. abdominalis. Molecular Ecology*, **23**, 5552–5565.
472
473 Cowman, P.F. & Bellwood, D.R. (2013) The historical biogeography of coral reef fishes: global
474 patterns of origination and dispersal. *Journal of Biogeography*, **40**, 209-224.
475
476 Craig, M.T., Eble, J.A., Bowen, B.W. & Robertson, D.R. (2007) High genetic connectivity across the
477 Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). *Marine Ecology*
478 *Progress Series*, **334**, 245–254.
479
480 Craig, M.T., Eble, J.A. & Bowen, B.W. (2010) Origins, ages, and population histories: Comparative
481 phylogeography of endemic Hawaiian butterflyfishes (genus *Chaetodon*). *Journal of Biogeography*, **37**,
482 2125 – 2136.
483
484 Delrieu-Trottin, E., Maynard, J., Planes, S. (2014) Endemic and widespread coral reef fishes have
485 similar mitochondrial genetic diversity. *Proceeding of the Royal Society B: Biological Sciences*, **281**.
486 doi: 10.1098/rspb.2014.1068.

DiBattista, J.D., Wilcox, C., Craig, M.T., Rocha, L.A. & Bowen, B.W. (2011) Phylogeography of the Pacific Blueline Surgeonfish *Acanthurus nigroris* reveals a cryptic 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., Rocha, L.A., Hobbs, J-P.A., He, S., Priest, M.A., Sinclair-Taylor, T.H., Bowen, B.W. & Berumen. M.L. (2015) When biogeographic provinces collide: Hybridization at the crossroads of three marine biogeographic provinces in the Arabian Sea. *Journal of Biogeography*, Online early

DiBattista J.D., Roberts, M, Baird, A.H., et al. (in review) A review of contemporary patterns of endemism for shallow water reef fauna in the Red Sea. *Journal of Biogeography*.

Drummond, A.J., Ashton, V., Buxton, V., Cheung, M., Cooper, V., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S. Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. & Wilson, A. (2010) Geneious version 5.4, Available from http://www.geneious.com.

511 Dwyer, G.S., Cronin, T.M., Baker, P.A., Raymo, M.E., Buzas, J.S. & Correge, T. (1995) North
512 Atlantic deepwater temperature change during late Pliocene and late Quaternary climatic cycles.
513 *Science*, **270**, 1347–1351.
514
515 Earl, D.A., & von Holdt, B.M. (2012) STRUCTURE HARVESTER: a website and program for
516 visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics*
517 *Resources*, **4**, 359-361.
518
519 Eble, J.A., Toonen, R.J. & Bowen, B.W. (2009) Endemism and dispersal: comparative phylogeography
520 of three surgeonfishes across the Hawaiian Archipelago. *Marine Biology*, **156**, 689-698.
521
522 Eble, J.A., Rocha, L.A., Craig, M.T. & Bowen, B.W. (2011) Not all larvae stay close to home: Long-
523 distance dispersal in Indo-Pacific reef fishes, with a focus on the Brown Surgeonfish (*Acanthurus*
524 *nigrofuscus*). *Journal of Marine Biology*, Article ID 518516.
525
526 Eble, J.A., Bowen, B.W. & Bernardi. G. (2015) Phylogeography of coral reef fishes. *Ecology of Fishes*
527 *on Coral Reefs* (ed. C. Mora). Pp. 64 – 75. University of Hawaii Press, Honolulu
528
529 Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the
530 software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
531
532 Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from
533 metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.
534 Genetics, 131, 479–491.

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

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

Fessler, J.L. & Westneat, M.W. (2007) Molecular phylogenetics of the butterflyfishes (Chaetodontidae): Taxonomy and biogeography of a global coral reef fish family. *Molecular Phylogenetics and Evolution*, **45**, 50–68.

Frankham, R. (1998) Inbreeding and extinction: island populations. *Conservation Biology*, **12**, 665–675.

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

Gaither, M.R., Toonen, R.J., Robertson, D.R., Planes, S. & Bowen, B.W. (2010) Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira*) and (*Lutjanus fulvus*). *Journal of Biogeography*, **37**, 133–147.

558 Gaither, M.R., Bowen, B.W., Bordenave, T.R., Rocha, L.A., Newman, S.J., Gomez, J.A., van
559 Herwerden, L. & Craig, M.T. (2011a) Phylogeography of the reef fish *Cephalopholis argus*
560 (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific barrier with contemporary
561 overlap in the coral triangle. *BMC Evolutionary Biology*, **11**, 189.
562
563 Gaither, M.R., Jones, S.A., Kelley, C., Newman, S.J., Sorenson, L., & Bowen, B.W. (2011b) High
564 connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific
565 with isolation of the Hawaiian Archipelago. *PLoS ONE*, **6**, e28913.
566
567 Gaither, M.R & Rocha, L.A. (2013) Origins of species richness in the Indo-Malay-Philippine
568 biodiversity hotspot: evidence for the centre of overlap hypothesis. *Journal of Biogeography*, **40**, 1638–
569 1648. doi:10.1111/jbi.12126.
570
571 Hasegawa, M., Kishino, H. & Yano, T. (1985) Dating of the human-ape splitting by a molecular clock
572 of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160-174.
573
574 Hobbs, J.-P.A., Frisch, A.J., Allen, G.R. & van Herwerden, L. (2009) Marine hybrid hotspot at Indo-
575 Pacific biogeographic border. *Biology Letters*, **5**, 258–61.
576
577 Hobbs, J.-P.A., Jones, G.P. & Munday, P.L. (2011) Extinction risk in endemic marine fishes.
578 *Conservation Biology*, **25**, 1053–1055. doi:10.1111/j.1523-1739.2011.01698.x.
579

580 Hobbs, J.-P.A., van Herwerden, L., Jerry, D.R., Jones, G.P. & Munday, P.L. (2013) High genetic
581 diversity in geographically remote populations of endemic and widespread coral reef angelfishes
582 (genus: Centropyge). *Diversity*, **5**, 39-50.
583
584 Hodge, J.R., van Herwerden. L. & Bellwood. D.R. (2014) Temporal evolution of coral reef fishes:
585 global patterns and disparity in isolated locations. *Journal of Biogeography*, **41**, 2115–2127.
586
587 Hourigan, T.F. & Reese, E.S. (1987) Mid-ocean isolation and the evolution of Hawaiian reef fishes.
588 *Trends in Ecology and Evolution*, **2**, 187-191.
589
590 Hsu, K.C., Chen, J.P. & Shao, K.T. (2007) Molecular phylogeny of *Chaetodon* (Teleostei:
591 Chaetodontidae) in the Indo-West Pacific: evolution in geminate species pairs and species groups. *The*
592 *Raffles Bulletin of Zoology Supplement* **14**, 77-86.
593
594 Kemp, J.M. (1998) Zoogeography of the coral reef fishes of the Socotra Archipelago. *Journal of*
595 *Biogeography*, **25**, 919-933.
596
597 Kobayashi, D.R. (2006) Colonization of the Hawaiian Archipelago via Johnston Atoll: a
598 characterization of oceanographic transport corridors for pelagic larvae using computer simulation.
599 *Coral Reefs*, **25**, 407-417.
600
601 Kuiter, R.H. (2002) Butterflyfishes, Bannerfishes and their Relatives. A comprehensive Guide
602 to Chaetodontidae and Microcanthidae. The Marine Fish Families Series, TMC Publishing,
603 Chorleywood, UK.

Lawton, R.J., Bay, L.K. & Pratchett, M.S. (2010) Isolation and characterization of 29 microsatellite loci for studies of population connectivity in the butterflyfishes *Chaetodon trifascialis* and *Chaetodon lunulatus*. *Conservation Genetics Resources,* **2**, 209-213.

Lawton, R.J., Messmer, V., Pratchett, M.S. & Bay, L.K. (2011) High gene flow across large geographic scales reduces extinction risk for a highly specialised coral feeding butterflyfish. *Molecular Ecology*, **20**, 3584-3598.

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

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

McMillan, W.O., Weight, 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.

627 Meirmans, P.G. & van Tienderen, P.H. (2004) GENOTYPE and GENODIVE: Two programs for the
628 analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4: 92-794.
629
630 Montanari, S.R., van Herwerden, L., Pratchett, M.S., Hobbs, J.P.A & Fugedi, A. (2012) Reef fish
631 hybridization: lessons learnt from butterflyfishes (genus *Chaetodon*). *Ecology and Evolution*, **2**, 310-
632 328.
633
634 Montanari, S.R., Hobbs, J-P.A., Pratchett, M.S., Bay, L.K. & van Herwerden, L. (2014) Does genetic
635 distance between parental species influence outcomes of hybridisation among coral reef
636 butterflyfishes? *Molecular Ecology*, **23**, 2757-2770.
637
638 Posada, D. (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, **25**,
639 253-1256.
640
641 Pratchett, M.S., Hoey, A.S., Feary, D.A., Bauman, A.G., Burt, J.A. & Riegl, B.M. (2013) Functional
642 composition of *Chaetodon* butterflyfishes at a peripheral and extreme coral reef location, the Persian
643 Gulf. *Marine Pollution Bulletin*, **72**, 333–341.
644
645 Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus
646 genotype data. *Genetics*, **155**, 945–959.
647
648 Randall, J.E. (1994) Twenty-two new records of fishes from the Red Sea. *Fauna Saudi Arabia*, **14**,
649 259–275.
650

651 Randall, J.E. (1998) Zoography of shore fishes of the Indo-Pacific region. *Zoological Studies*, **37**, 227-
652 268.
653
654 Randall, J.E. (2007) *Reef and Shore Fishes of the Hawaiian Islands*. University of Hawaii Press,
655 Honolulu.
656
657 Raymond, M. & Rousset, F. (1995) GENEPOP (version 1.2): Population genetics software for exact
658 tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.
659
660 Reece, J.S., Bowen, B.W. & Larson, A.F. (2011) Long larval duration in moray eels (Muraenidae)
661 ensures ocean-wide connectivity despite differences in adult niche breadth. *Marine Ecology Progress*
662 *Series*, **437**, 269–277.
663
664 Righton, D., Kemp, J. & Ormond, R. (1996) Biogeography, community structure and diversity of Red
665 Sea and western Indian Ocean butterflyfishes. *Journal of the Marine Biological Association of the*
666 *United Kingdom*, **76,** 223–228.
667
668 Roberts, C.M., Shepherd, A.R.D. & Ormond, R.F.G. (1992) Large scale variation in assemblage
669 structure of Red Sea butterflyfishes and angelfishes. *Journal of Biogeography*, **19**, 239-250.
670
671 Rocha, L.A., Craig, M.T. & Bowen, B.W. (2007) Phylogeography and the conservation genetics of
672 coral reef fishes. *Coral Reefs*, **26**, 501-512.
673

674 Selkoe, K.A. & Toonen, R.J. (2006) Microsatellites for ecologists: a practical guide to using and
675 evaluating microsatellite markers. *Ecology Letters*, **9**, 615-629.
676
677 Seutin, G., White, B.N. & Boag, P.T. (1991) Preservation of avian blood and tissue samples for DNA
678 analyses. *Canadian Journal of Zoology*, **69**, 82-90.
679
680 Sheppard, C.R.C. (1998) Biodiversity patterns in Indian Ocean corals, and effects of taxonomic error in
681 data. *Biodiversity and Conservation*, **7**, 847-868.
682
683 Sheppard, C., Al-Husiani, M., Al-Jamali, F. & *et al*. (2010) The Gulf: A young sea in decline. *Marine*
684 *Pollution Bulletin*, **60**, 13–38.
685
686 Sheppard, C., Ateweberhan, M., Bowen, B.W. *et al*. (2012) Reefs and islands of the Chagos
687 Archipelago, Indian Ocean: Why it is the world's largest no-take marine protected area. Aquatic
688 Conservation: Marine and Freshwater Ecosystems, **22**, 232-261.
689
690 Soeparno, Nakamura, Y., Shibuno, T. & Yamaoka, K. (2012) Relationship between pelagic larval
691 duration and abundance of tropical fishes on temperate coasts of Japan. *Journal of Fish Biology*, **80**,
692 346–357.
693
694 Swofford, D.L. (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).
695 Version 4. Sinauer Associates, Sunderland, Massachusetts.
696

697 Szabo, Z., Snelgrove, B., Craig, M.T., Rocha, L.A. & Bowen, B.W. 2014. Phylogeography of the
698 Manybar Goatfish, *Parupeneus multifasciatus* reveals moderate structure between the Central and
699 North Pacific and a cryptic endemic species in the Marquesas. *Bulletin of Marine Science*, **90**, 493 –
700 512.
701
702 Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region
703 of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
704
705 Tamura, K., Peterson, D., Peterson, N., Stecher, G. , Nei, M. & Kumar, S. (2011) MEGA5: Molecular
706 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum
707 parsimony methods. Molecular Biology and Evolution, **28**, 2731-2739.
708
709 Toonen, R.J., Andrews, K.R., Baums, I.B., *et al*. (2011) Defining boundaries for ecosystem-based
710 management: a multispecies case study of marine connectivity across the Hawaiian archipelago.
711 *Journal of Marine Biology*, Article ID 460173.
712
713 van Oosterhout, C., Hutchinson, W.F., Willis, D.P.M. & Shipley, P. (2004) MICRO-CHECKER:
714 software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology*
715 *Notes*, **4**, 535–538.
716
717 Vogler, C., Benzie, J.A.H., Tenggardjaja, K., Ambariyanto, Barber, P.H. & Wörheide, G. (2012)
718 Phylogeography of the crown-of-thorns starfish in the Indian Ocean. *Coral Reefs*, **32**, 515–525.
719

Waldrop, E. 2014. Phylogeography and Evolution of Butterflyfishes in the Subgenus Corallochaetodon: *Chaetodon lunulatus, Chaaetodon trifasciatus, Chaetodon austriacus, Chaetodon melapterus*. Thesis, University of Hawaii, Honolulu

Williams, D.F., Peck, J., Karabanov, E.B., Prokopenko, A.A., Kravchinsky, V., King, J. & Kuzmin, M.I. (1997) Lake Baikal record of continental climate response to orbital insolation during the past five million years. *Science*, **278**, 1114-1117.

Zekeria, Z.A., Afeworki, Y. & Videler, J.J. (2005) The distribution patterns of Red Sea Chaetodontid assemblages. Aquatic Conservation: *Marine and Freshwater Ecosystems,* **15**, S71–S76.

SUPPORTING INFORMATION

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

Appendix S1: Additional materials & methods

Appendix S2: Supporting tables & figures

BIOSKETCH

Lead author Ellen Waldrop conducted this research as a M.Sc. thesis project at the University of Hawai‘i. The authors are interested in the origins of marine biodiversity and the prudent management of evolutionary lineages.

Author contributions: B.W.B. initiated the research; E.W., J.-P.A.H., J.D.D., L.A.R., R.K.K., M.L.B. and B.W.B. conducted field expeditions and sampling; E.W and J.D.D. provided genetic data; E.W.

743 analysed the data; E.W., B.W.B., J.P.H. and J.D.D contributed to the writing; and all authors
744 commented on the final draft.
745
746 Editor: Michelle Gaither
747
748

749 **Table 1**. Sample size and molecular diversity indices for *Chaetodon lunulatus, C. trifasciatus, C.*
750 *melapterus* and *C. austriacus* based on mtDNA cytochrome *b* sequence data (significant Fu's *Fs* values
751 are in bold, $P < 0.02$). For *C. trifasciatus*, specimens from the eastern Indian Ocean (Cocos-Keeling
752 Islands and adjacent Christmas Island) were pooled to increase statistical power as they were
753 indistinguishable in preliminary analyses.

Location	***N***	**Number of Haplotypes**	**Haplotype diversity (*h* ± SD)**			**Nucleotide diversity (*π* ± SD)**			**Fu's *Fs***
C. lunulatus									
Christmas Island	6	4	0.867	+/-	0.129	0.005	+/-	0.004	0.24
American Samoa	15	5	0.714	+/-	0.081	0.005	+/-	0.003	1.40
Fiji	30	10	0.602	+/-	0.104	0.004	+/-	0.003	-1.92
Kanton Island	15	5	0.695	+/-	0.109	0.004	+/-	0.003	0.95
Marshall Islands	29	8	0.727	+/-	0.057	0.005	+/-	0.003	0.91
Mo‘orea	32	8	0.669	+/-	0.086	0.005	+/-	0.003	-0.04
Okinawa	8	4	0.643	+/-	0.184	0.004	+/-	0.003	0.73
Pohnpei	30	10	0.782	+/-	0.065	0.005	+/-	0.003	-0.57
Kiribati	22	3	0.589	+/-	0.066	0.004	+/-	0.003	4.63
Palau	26	2	0.471	+/-	0.063	0.004	+/-	0.002	6.68
Johnston Atoll	31	2	0.516	+/-	0.024	0.004	+/-	0.003	7.63
MHI	33	2	0.504	+/-	0.034	0.004	+/-	0.003	7.64
NWHI	161	13	0.452	+/-	0.048	0.001	+/-	0.001	-0.51
C. trifasciatus									
Diego Garcia	29	8	0.672	+/-	0.074	0.001	+/-	0.001	**-4.538**
Seychelles	21	9	0.795	+/-	0.077	0.088	+/-	0.044	9.843
Christmas Island	14	7	0.802	+/-	0.094	0.010	+/-	0.006	0.959
Indonesia	5	3	0.700	+/-	0.218	0.002	+/-	0.002	0.061
C. melapterus									
Maskali	17	5	0.353	+/-	0.353	0.001	+/-	0.001	**-2.527**
Obock	29	7	0.778	+/-	0.584	0.001	+/-	0.001	**-3.754**
Bay of Ghoubbet	15	1	0.000	+/-	0.000	0.000	+/-	0.000	na
Oman	34	9	0.631	+/-	0.507	0.001	+/-	0.001	**-7.615**
C. austriacus									
Al Lith	10	2	0.200	+/-	0.154	0.000	+/-	0.000	na
Jazirat Baraqan	10	6	0.844	+/-	0.103	0.002	+/-	0.002	**-3.127**
Yanbu	10	7	0.866	+/-	0.107	0.001	+/-	0.001	**-1.404**

754

755 **Table 2.** Matrix of population pairwise Φ_{ST} values (above diagonal) and associated *P* values (below diagonal) based on 605 bp of mtDNA
756 cytochrome *b* sequence data from *Chaetodon lunulatus.* Significant *P* values are indicated in bold ($P < 0.05$). All negative Φ_{ST} values were
757 adjusted to 0.

Location	Christmas Island	American Samoa	Fiji	Kanton Island	Marshall Island	Mo‘orea	Okinawa	Pohnpei	Kiribati	Palau	Johnston Atoll	MHI	NWHI
Christmas Island	—	0	0.097	0.012	0	**0.284**	0.107	0	0	0.084	0.006	0.003	**0.597**
American Samoa	0.568	—	**0.105**	0.095	0	**0.286**	**0.074**	0	0	0.040	0	0	**0.507**
Fiji	0.108	**0.036**	—	0	**0.086**	**0.478**	0	0.024	0.022	0.000	**0.083**	**0.162**	**0.114**
Kanton Island	0.333	0.081	0.477	—	0.079	**0.470**	0	0	0.031	0.040	0.105	**0.178**	**0.245**
Marshall Islands	0.414	0.973	**0.036**	0.099	—	**0.307**	0.050	0	0	0.023	0	0	**0.431**
Mo‘orea	**0.036**	**<0.001**	**0.000**	**0.000**	**0.000**	—	**0.463**	**0.370**	**0.371**	**0.431**	**0.342**	**0.298**	**0.757**
Okinawa	0.234	**0.036**	0.847	0.387	0.189	**<0.001**	—	0.008	0	0	0.037	0.125	**0.099**
Pohnpei	0.658	0.387	0.144	0.423	0.369	**<0.001**	0.252	—	0	0.010	0.016	**0.055**	**0.332**
Kiribati	0.324	0.514	0.126	0.216	0.640	**<0.001**	0.306	0.667	—	0	0	0.017	**0.335**
Palau	0.252	0.198	0.324	0.126	0.234	**<0.001**	0.396	0.207	0.559	—	0.003	0.068	**0.228**
Johnston Atoll	0.324	0.450	**0.018**	0.063	0.577	**<0.001**	0.108	0.189	0.631	0.432	—	0	**0.405**
MHI	0.279	0.550	**0.009**	**0.018**	0.423	**<0.001**	0.099	**0.045**	0.342	0.108	0.622	—	**0.509**
NWHI	**0.009**	**<0.001**	**0.009**	**<0.001**	**<0.001**	**<0.001**	**0.018**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	—

758

759 **Table 3.** Matrix of population pairwise Φ_{ST} values (above diagonal) and associated *P* values (below
760 diagonal) based on 605 bp of mtDNA cytochrome *b* sequence data from *Chaetodon trifasciatus*, *C.*
761 *melapterus* and *C. austriacus*. All negative Φ_{ST} values were adjusted to 0.

C. trifasciatus				
Location	Diego Garcia	Seychelles	Christmas Island	Indonesia
Diego Garcia	–	0.014	0.027	0
Seychelles	0.268	–	0	0
Christmas Island	0.238	0.961	–	0
Indonesia	0.483	0.769	0.678	–
C. melapterus				
Location	Maskali	Obock	Bay of Ghoubbet	Oman
Maskali	–	0.030	0	0.001
Obock	0.108	–	0.022	0.007
Bay of Ghoubbet	0.991	0.270	–	0
Oman	0.459	0.288	0.667	–
C. austriacus				
Location	Al Lith	Jazirat Baraqan	Yanbu	
Al Lith	–	0.095	0.028	
Jazirat Baraqan	0.207	–	0	
Yanbu	0.491	0.573	–	

762

763 **Table 4.** Matrix of population pairwise F_{ST} values (above diagonal) and associated *P* values (below diagonal) based on microsatellite
764 genotypes for *Chaetodon lunulatus.* Significant *P* values are highlighted in bold ($P < 0.05$). All negative F_{ST} values were adjusted to 0.

Location	Christmas Island	Indonesia	American Samoa	Fiji	Kanton Island	Marshall Islands	Mo‘orea	Okinawa	Pohnpei	Kiribati	Palau	Johnston Atoll	MHI	NWHI
Christmas Island	—	0	0.003	0.001	0.012	0.006	**0.041**	0.010	0.006	0	0.011	**0.084**	**0.032**	**0.090**
Indonesia	0.498	—	0.007	0.002	0.001	0	**0.030**	0.002	0.0	0	0	**0.079**	**0.024**	**0.078**
American Samoa	0.378	0.067	—	**0.009**	0.002	0.006	**0.027**	0.012	**0.010**	0	0.007	**0.082**	**0.037**	**0.075**
Fiji	0.396	0.267	**0.036**	—	0.002	0.002	**0.030**	0.007	0.005	0.000	**0.007**	**0.088**	**0.030**	**0.089**
Kanton Island	0.124	0.411	0.322	0.260	—	0	**0.023**	0.003	0.001	0	0.004	**0.087**	**0.035**	**0.076**
Marshall Islands	0.217	0.706	0.067	0.150	0.772	—	**0.029**	0.005	0.000	0.000	0.002	**0.084**	**0.030**	**0.079**
Mo‘orea	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	—	0.056	0.032	0.024	0.029	**0.087**	**0.058**	**0.095**
Okinawa	0.203	0.300	0.089	0.093	0.331	0.116	**<0.001**	—	0.005	0.007	0.005	**0.096**	**0.034**	**0.082**
Pohnpei	0.232	0.676	**0.022**	0.071	0.361	0.531	**<0.001**	0.151	—	0	0	**0.085**	**0.029**	**0.081**
Kiribati	0.497	0.744	0.602	0.443	0.779	0.394	**<0.001**	0.109	0.773	—	0	**0.076**	**0.023**	**0.067**
Palau	0.128	0.779	0.072	**0.017**	0.154	0.203	**<0.001**	0.140	0.441	0.554	—	**0.080**	**0.023**	**0.078**
Johnston	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	—	**0.051**	**0.038**
MHI	**0.005**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	—	**0.053**
NWHI	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	—

765

Table 5. Matrix of population pairwise F_{ST} values (above diagonal) and associated *P* values (below diagonal) based on microsatellite genotypes for *Chaetodon trifasciatus*. Significant *P* values are highlighted in bold ($P < 0.05$). All negative F_{ST} values were adjusted to 0.

Location	Diego Garcia	Seychelles	Christmas Island	Indonesia
Diego Garcia	–	0.005	0.006	0.012
Seychelles	**0.047**	–	0	0
Christmas Island	**0.013**	0.742	–	0.001
Indonesia	**0.018**	0.496	0.350	–

TITLES AND LEGENDS TO FIGURES

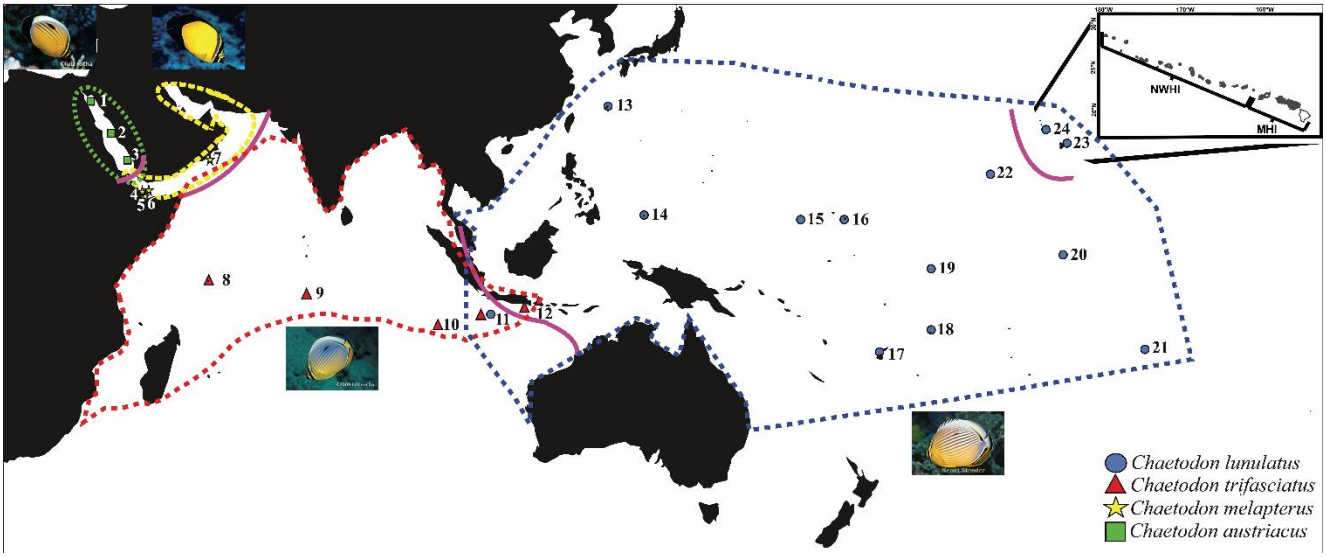
Figure 1. Distribution map of the *corallochaetodon* subgenus (redrawn from Blum, 1989). *Chaetodon lunulatus* (blue, widespread Pacific Ocean), *C. trifasciatus* (red, widespread Indian Ocean), *C. austricaus* (green, largely restricted to the northern and central Red Sea; but see DiBattista *et al.*, in review) and *C. melapterus* (yellow, restricted to the southern Red Sea through the Arabian Gulf). The known geographic range of each species is outlined with a dotted line and solid pink lines represent known marine biogeographic barriers (Hsu *et al*., 2007) that influence the genetic partitions and evolution of *corallochaetodon.* Sample locations are shown with species-specific coloured symbols and numbers that correspond to the following location names: 1. Jazirat Baraqan, 2. Yanbu, 3. Al Lith, 4. Obock, 5. Bay of Ghoubbet, 6. Maskali, 7. Oman, 8. Seychelles, 9. Diego Garcia, 10. Cocos (Keeling) Islands, 11. Christmas Island, 12. Indonesia, 13. Okinawa, 14. Palau, 15. Pohnpei, 16. Marshall Islands, 17. Fiji, 18. American Samoa, 19. Kanton Island, 20. Kiribati, 21. Mo‘orea, 22. Johnston Atoll, 23. Main Hawaiian Islands, 24. Northwestern Hawaiian Islands. Sample sizes for each location are presented in Table 1. Photo Credits: L.A. Rocha for *C. austriacus* and *C. trifasciatus,* Keoki Stender for *C. lunulatus*.

Figure 2. Neighbour-joining tree based on mtDNA cytochrome *b* sequences, highlighting the relationship between sister species in the subgenus *corallochaetodon* (bootstrap values shown based on 1000 replicates). For simplicity, only a representative subset of specimens is shown. Maximum-likelihood and maximum-parsimony trees yielded the same topology among species. *Chaetodon vagabundus* is used as an outgroup (Genbank accession number JF458006). Abbreviations: *C. lunulatus* = Clu, *C. trifasciatus* = Ctt, *C. melapterus* = Cml and *C. austriacus* = Cau.

Figure 3. Statistical parsimony network for *Chaetodon lunulatus* (pink, purple, blue shades), *C. trifasciatus* (green shades), *C. melapterus* (yellow and orange) and *C. austriacus* (red) based on mtDNA cytochrome *b* sequences. The area of each circle is proportional to the abundance of the respective haplotype: small circles indicate rare or unique haplotypes and the largest circle indicate the most common haplotype observed in 286 sampled individuals. Black bars and black branches represent a single mutation (unless otherwise noted) and colours indicate haplotype sampling location (see key).

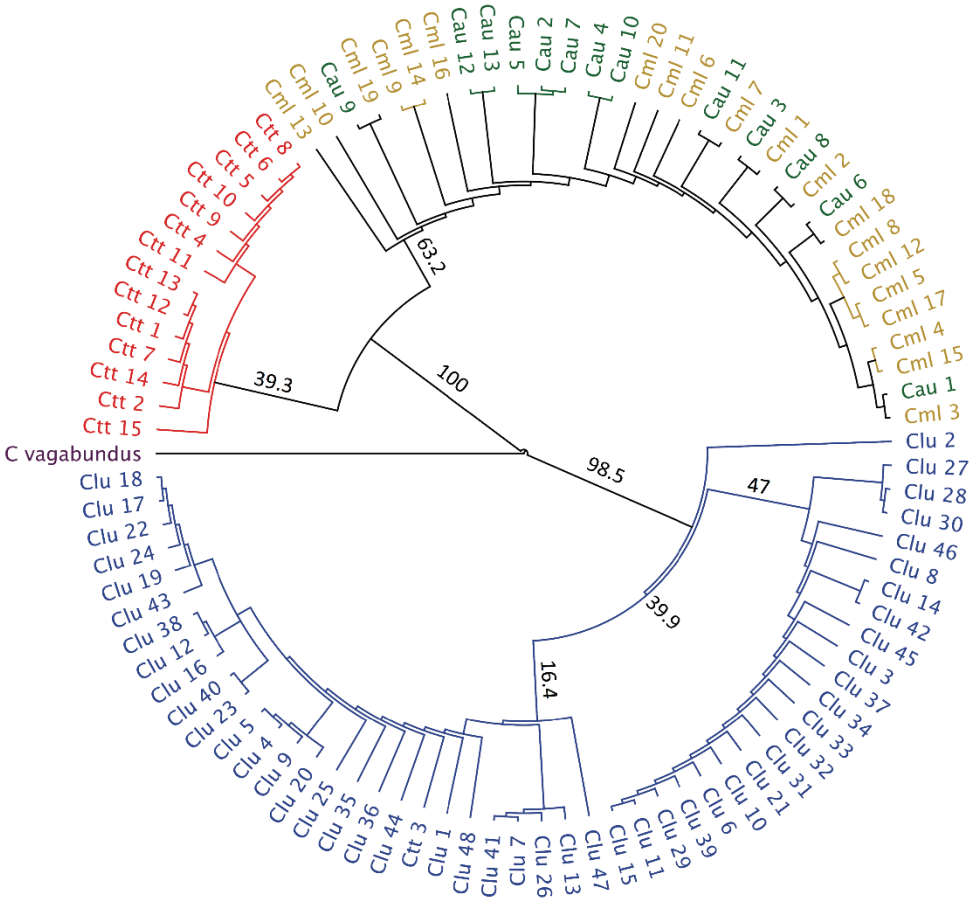
Figure 4. STRUCTURE bar plot for *Chaetodon lunulatus* showing the highest mean probability of K = 3. Locations: 1. Christmas Island, 2. Indonesia, 3. Palau, 4. Okinawa, 5. Pohnpei, 6. Marshall Islands, 7. Fiji, 8. American Samoa, 9. Mo‘orea, 10. Kanton Island, 11. Kiribati, 12. Johnston Atoll, 13. MHI, 14. NWHI.

Figure 5. STRUCTURE bar plot for *Chaetodon trifasciatus*, showing the highest mean probability of K = 2. Locations: 1. Diego Garcia, 2. Seychelles, 3. Christmas Island, 4. Indonesia.



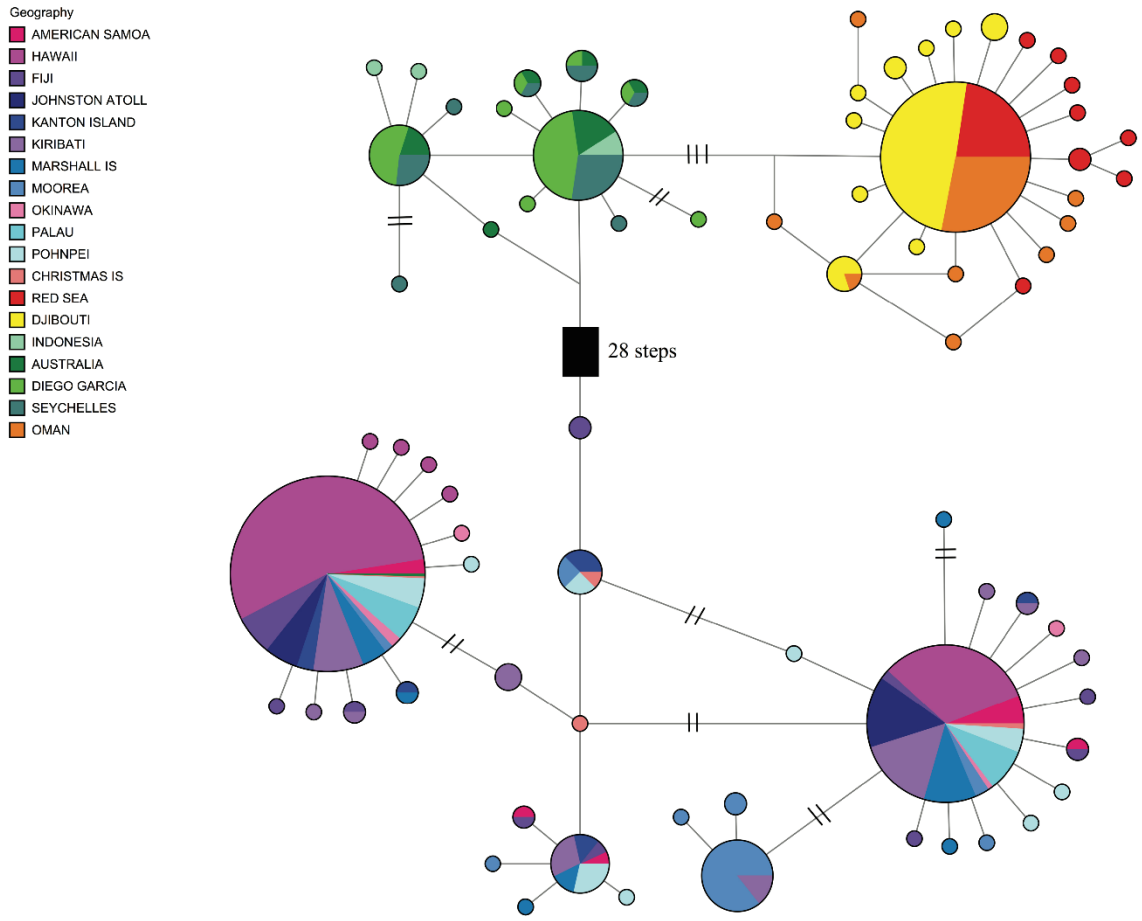
810

811



812
813

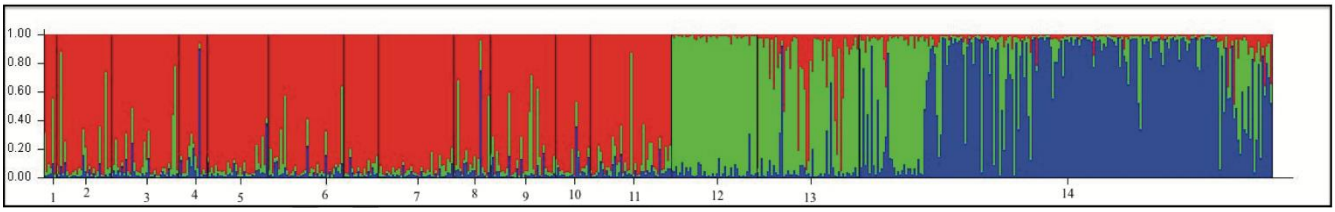
814



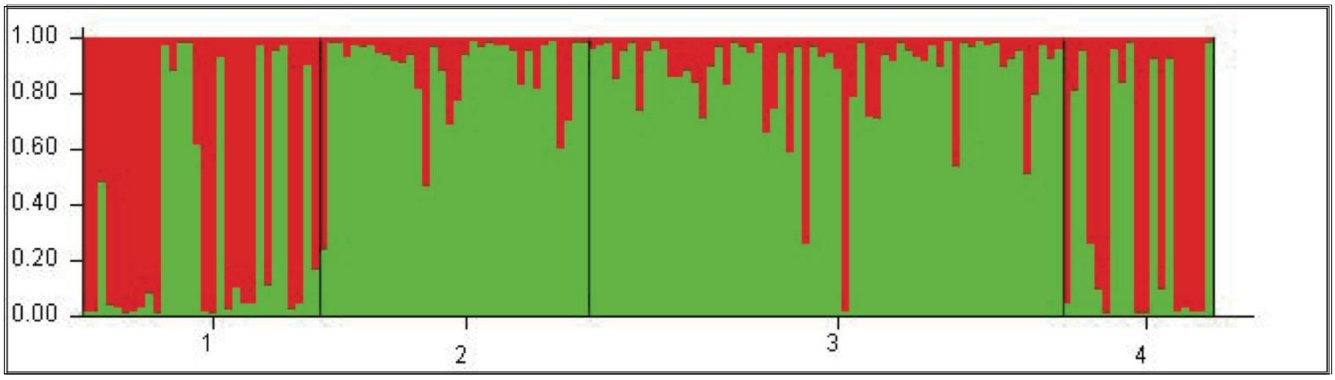
815
816

817

818



819



820

821