This is the peer reviewed version of the following article: "Di Battista, J. and Waldrop, E. and Rocha, L. and Craig, M. and Berumen, M. and Bowen, B. 2015. Blinded by the bright: A lack of congruence between colour morphs, phylogeography and taxonomy for a cosmopolitan Indo-Pacific butterflyfish, Chaetodon auriga. Journal of Biogeography. 42 (10): pp. 1919-1929.", which has been published in final form at doi:10.1111/jbi.12572. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving at http://olabout.wiley.com/WileyCDA/Section/id-828039.html

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

2 Original Article

3 LRH: J. D. DiBattista *et al.*

4 RRH: Colour patterns and phylogeography of *Chaetodon auriga*

5

6 Blinded by the bright: A lack of congruence between colour morphs, phylogeography and 7 taxonomy for a cosmopolitan Indo-Pacific butterflyfish, *Chaetodon auriga*

8

9 Joseph D. DiBattista[1,2*], Ellen Waldrop[3], Luiz A. Rocha[4], Matthew T. Craig[5], Michael L. Berumen[1],

10 Brian W. Bowen[3]

11

12 *[1]Division of Biological and Environmental Science and Engineering, King Abdullah University of*

13 *Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia, [2]Department of Environment*

14 *and Agriculture, Curtin University, PO Box U1987, Perth, WA 6845, Australia, [3]Hawai'i Institute of*

15 *Marine Biology, Kāne'ohe, HI 96744, USA, [4]Section of Ichthyology, California Academy of Sciences,*

16 *San Francisco, CA 94118, USA, [5]Southwest Fisheries Science Center, NOAA National Marine Fisheries*

17 *Service, La Jolla, CA 92037, USA*

18

19

20 *Correspondence: Joseph D. DiBattista, Department of Environment and Agriculture, Curtin University,

21 PO Box U1987, Perth, WA 6845, Australia

22 E-mail: josephdibattista@gmail.com

23

24 Word Count: 6,953

25

26 **ABSTRACT**

27 **Aim** We assess genetic differentiation among biogeographical provinces and colour morphs of the
28 threadfin butterflyfish, *Chaetodon auriga*. This species is among the most broadly distributed
29 butterflyfishes in the world, occurring on reefs from the Red Sea and western Indian Ocean to French
30 Polynesia and Hawai'i. The Red Sea form lacks a conspicuous "eye-spot" on the dorsal fin, which may
31 indicate an evolutionary distinction.

32

33 **Location** Red Sea, Indian Ocean and Pacific Ocean.

34

35 **Methods** Specimens were obtained at 17 locations (N = 358) spanning the entire range of this species.
36 Genetic data include 669 base pairs of mitochondrial DNA (mtDNA) cytochrome *b* and allele
37 frequencies at six microsatellite loci. Analysis of molecular variance, STRUCTURE plots, haplotype
38 networks and estimates of population expansion time were used to assess phylogeographical patterns.

39

40 **Results** Population structure was low overall, but significant and concordant between molecular markers
41 (mtDNA: $\Phi_{ST} = 0.027$, $P < 0.001$; microsatellites: $F_{ST} = 0.023$, $P < 0.001$). Significant population-level
42 partitions were only detected at peripheral locations including the Red Sea and Hawai'i. Populations in
43 the Red Sea and Socotra are older (111,940 to 223,881 years) relative to all other sites (16,343 to 87,910
44 years).

45

46 **Main conclusions** We find little genetic evidence to support an evolutionary partition of a previously

47 proposed Red Sea subspecies. The oldest estimate of population expansion in the Red Sea and adjacent
48 Gulf of Aden indicates a putative refuge in this region during Pleistocene glacial cycles. The finding of
49 population separations at the limits of the range, in the Red Sea and Hawai‘i, is consistent with
50 peripheral speciation.
51
52 **Keywords**
53 **Coral reef fish, marine biogeography, microsatellite, mitochondrial DNA, population expansion**
54 **time, subspecies**
55
56 **INTRODUCTION**
57 Colouration plays an important role in the taxonomic classification of reef fishes and is frequently the
58 sole character used to distinguish closely related species. Its evolutionary significance, however, is
59 uncertain (McMillan *et al.,* 1999; Bernardi *et al.,* 2002), since colour variation can be a result of
60 phenotypic plasticity rather than reproductive isolation (Grady & Quattro, 1999). In addition,
61 colouration may evolve faster than morphological and genetic characters (Schultz *et al.,* 2007).
62 Colour polymorphisms within the same species are relatively common (e.g. brown dottyback,
63 Messmer *et al.,* 2005) and several mechanisms have been proposed to explain their existence. Many reef
64 fishes are distinguished primarily by colour, yet colouration is not necessarily a species-specific
65 diagnostic character, particularly for widespread species (flame angelfish, Schultz *et al.,* 2007; King
66 Demoiselle, Drew *et al.,* 2008). In some cases there is greater concordance between genetics and
67 geography than between genetics and colouration (McMillan & Palumbi, 1995; DiBattista *et al.,* 2012a).
68 Like many reef fishes, the butterflyfishes (family: Chaetodontidae) have a spectacular variety of
69 colour patterns, and there appears to be a link between species diversification and colour variation

70 (Blum, 1989). The significance of colour in this group, however, must be interpreted with caution. For
71 example, McMillan *et al.* (1999) observed colour pattern evolution associated with genetic divergence in
72 *Chaetodon multicinctus*, a Hawaiian endemic, but not in its two sister species (*C. punctatofasciatus* and
73 *C. pelewensis*) distributed across the Indo-West Pacific.

74 The threadfin butterflyfish, *Chaetodon auriga* Forsskål, 1775, is among the most widespread reef
75 fishes on the planet, occurring from Hawai‘i and French Polynesia to the Red Sea. One colour morph
76 restricted to the Red Sea (and almost exclusively in the northern and central regions) lacks the
77 conspicuous "eye-spot" or ocellum on the soft dorsal fin (Fig. 1). Consequently, the Red Sea population
78 was recognized as a subspecies (*C. auriga auriga*; Allen, 1979). All other individuals (those with the
79 dark spot), including those in Socotra and Oman just outside the Red Sea, are assigned to *C. auriga*
80 *setifer* (Allen, 1979). This distinction is notable because the threadfin butterfly has a relatively high
81 dispersal potential, with a pelagic larval duration (PLD) of 40 to 53 days (Leis, 1989). It is also
82 interesting given that isolated peripheral reef habitats (like the Red Sea) may be sources of evolutionary
83 novelty and contribute marine biodiversity to the broader Indo-West Pacific (Bowen *et al.,* 2013).

84 Preliminary genetic comparisons between *C. auriga* in the Red Sea and Western Indian Ocean
85 (WIO) detected mtDNA haplotype frequency differences between these two regions, although they were
86 only marginally significant (DiBattista *et al.*, 2013). The Red Sea also had low genetic diversity
87 compared to WIO sites (e.g. Seychelles and Diego Garcia). Based on these findings, our goals are to:

88 1) Characterize genetic diversity across the species range to resolve demographic histories, with
89 particular emphasis on the Red Sea;
90 2) Define population genetic structure across the range to resolve the relationship between genetic
91 partitions and biogeographical provinces.

92

93 **MATERIALS AND METHODS**

94 **Sample collection**

95 We collected 358 *C. auriga* tissue samples (fin clip or gills) at 17 sites while scuba diving or snorkeling
96 between 2006 and 2011 (Fig. 1). Tissues were preserved in a saturated salt-DMSO solution, total
97 genomic DNA was extracted using a "HotSHOT" protocol (Meeker *et al.*, 2007) and samples
98 subsequently stored at -20 ˚C.

99

100 **Mitochondrial DNA sequencing**

101 A 669 base pair (bp) segment of the mtDNA cytochrome *b* (cyt *b*) gene was resolved using heavy-strand
102 (5' - GTGACTTGAAAAACCACCGTTG - 3'; Song *et al.*, 1998) and light-strand primers (5' -
103 AATAGGAAGTATCATTCGGGTTTGATG - 3'; Taberlet *et al.*, 1992). Polymerase chain reaction
104 (PCR) conditions and product visualization followed protocols described in DiBattista *et al*. (2013). All
105 samples were sequenced in the forward direction with fluorescent dye terminators (BigDye 3.1, Applied
106 Biosystems Inc., Foster City, CA, USA) and analyzed using an ABI 3130XL Genetic Analyzer (Applied
107 Biosystems). The sequences were aligned, edited and trimmed to a common length using Geneious Pro
108 4.8.4 (Drummond *et al.*, 2009); all cyt *b* sequences were deposited in GenBank (accession numbers:
109 KM488667 to KM488795). jModelTest 1.0.1 (Posada, 2008) was used with an Akaike information
110 criterion (*AIC*) test and the TrN model (Tamura & Nei, 1993) was selected for subsequent analyses.

111 ARLEQUIN 3.5.1.2 (Excoffier *et al.*, 2005) was used to calculate haplotype (*h*) and nucleotide
112 diversity (π), as well as to test for population structure. Genetic differentiation among sampling sites was
113 first estimated with analysis of molecular variance (AMOVA) based on pairwise comparisons of sample
114 groups; deviations from null distributions were tested with non-parametric permutation procedures (N =
115 99,999). Pairwise Φ_{ST} statistics were also generated in ARLEQUIN, significance tested by permutation

116 (N = 99,999) and P-values adjusted according to the modified false discovery rate (FDR) method
117 (Narum, 2006). Patterns of significant genetic differentiation were congruent between FDR and more
118 conservative methods of correction (i.e. Bonferonni; data not shown). Multiple sites within French
119 Polynesia, the Hawaiian Islands and the Red Sea were grouped based on preliminary findings of genetic
120 homogeneity. We used the Isolation by Distance (IBD) Web Service 3.23 to detect correlations between
121 geographic and genetic distances (Jensen *et al.,* 2005). To mitigate any false positives, we tested: 1) the
122 whole range, 2) the range minus Red Sea/Socotra and 3) the range minus Hawai‘i.

123 Evolutionary relationships among haplotypes were estimated with an unrooted statistical
124 parsimony network using NETWORK 4.5.1.0 (www.fluxus-engineering.com/network_terms.htm) with
125 a median joining algorithm and default settings (Bandelt *et al.,* 1999).

126 Deviations from neutrality were assessed with Fu's F_S (Fu, 1997) for each group using
127 ARLEQUIN; significance was tested with 99,999 permutations. Negative (and significant) F_S values
128 indicate recent population expansion or selection. Time since most recent population expansion was
129 estimated using the parameter τ for each group (Rogers & Harpending, 1992) by applying the equation τ
130 $= 2\mu t$, where t is the age of the population in generations and μ is the mutation rate per generation for the
131 sequence (μ = number of bp • divergence rate within a lineage • generation time in years). A range of cyt
132 *b* mutation rates are available from previous fish studies: 2% per Myr between lineages or 1% within
133 lineages (Bowen *et al.*, 2001) and 1.55% per Myr within lineages or 1.55×10^{-8} mutations per site per
134 year (Lessios, 2008). While generation time is unknown for our study species, we conservatively used
135 an estimate of 3 years based on age/size distributions for other butterflyfishes (Berumen, 2005; Craig *et*
136 *al.,* 2010). Given that our interest lies in rank order time since expansion, rather than the absolute time
137 values, these approximations should be precise enough to support our conclusions.
138

139 **Microsatellite genotyping and analysis**

140 Six microsatellite loci were chosen from the suite developed by Berumen *et al.* (2009) and Lawton *et al.*
141 (2010), and validated more broadly in Chaetodontidae (Lawton *et al.,* 2011). PCR conditions and
142 product visualization followed protocols described by Berumen *et al.* (2009). PCR products labeled with
143 different fluorescent dyes were pooled for genotyping at equimolar concentrations using an ABI
144 3130XL Genetic Analyzer (Applied Biosystems) along with a labeled internal size standards (LIZ-500;
145 Applied Biosystems). Allele sizes were assigned with the Geneious Pro 5.6.7. All markers reliably
146 amplified and product sizes were consistent with expectations (Berumen *et al.,* 2009; Lawton *et al.,*
147 2010; Lawton *et al.,* 2011; Montanari *et al.,* 2012). A few sampling sites were not included for
148 microsatellite analysis because of small sample size (Durban, $N = 2$; Johnston Atoll, $N = 1$; Madagascar,
149 $N = 8$; Zanzibar, $N = 2$) or inconsistent amplification (Socotra, $N = 15$).

150 For each locus the mean number of alleles (N_{A}), observed (H_O) and expected (H_E)
151 heterozygosities, Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were assessed
152 with GENEPOP 4.2.2 and ARLEQUIN. Allelic richness was determined with FSTAT 2.9.3.2.
153 Significance levels for multiple comparisons were adjusted using false discovery rate method (Narum,
154 2006). MICRO-CHECKER 2.2.3 was used to identify genotyping errors including null alleles, allelic
155 dropouts and stutter peaks (Van Oosterhout *et al.,* 2004); significance levels for multiple comparisons
156 were adjusted using the sequential Bonferonni correction with default settings. Overall population
157 structure and pairwise comparisons (F_{ST} calculations) were estimated with ARLEQUIN. To facilitate
158 comparisons to other studies, an additional diversity measure, Jost's D (Jost, 2008), was estimated using
159 SPADE (Chao *et al.,* 2008). This metric compensates for the downward bias in F_{ST} produced by within-
160 population heterozygosity, a recurring problem with microsatellite markers (Bird *et al.,* 2011). IBD tests
161 were further conducted on the microsatellite dataset as outlined above.

162 STRUCTURE 2.3.2 was used to assign individuals to genetic clusters (populations) without bias
163 from geographical locations (Pritchard *et al.,* 2000). STRUCTURE uses a Bayesian approach to assign
164 individual multi-locus genotypes to clusters (K) by minimising deviations from Hardy-Weinberg and
165 linkage equilibrium. The most likely number of clusters was identified by testing the probability of K =1
166 to K = 12. Analyses were repeated five times and the results averaged. Each run consisted of 1,000,000
167 MCMC repetitions, a burn-in of 10,000 iterations and correlated allele frequencies and admixed
168 populations were assumed (as per DiBattista *et al.,* 2012b). STRUCTURE HARVESTER 0.6.94
169 identified the most likely K value (genetic groups) (Evanno *et al.,* 2005; Earl & vonHoldt, 2012).
170 A discriminant analysis of principal components (DAPC; Jombart *et al.,* 2010) was also run on
171 all loci to investigate the relationship between genotype and geographical location. The number of
172 principal components retained for genotypic variability was equal to the number of individuals divided
173 by three; the number of DA eigenvectors corresponded to the number of populations minus one.
174 Although different from the admixture coefficients of STRUCTURE, DAPC can still be interpreted as
175 proximities of individuals to different clusters based on the retained discriminant functions.
176
177 **RESULTS**
178 **Molecular characteristics**
179 Cyt *b* sequences from *C. auriga* included 33 haplotypes (4 to 10 within-sites), with haplotype and
180 nucleotide diversity ranging from h = 0.20 to 0.86 and π = 0.00031 to 0.00214 (Table 2). Haplotype and
181 nucleotide diversity was almost twice as high at all other sites compared to the Red Sea (h = 0.20 $\pm$ 0.08,
182 π = 0.00031 $\pm$ 0.00043) and French Polynesia (h = 0.34 $\pm$ 0.11, π = 0.00055 $\pm$ 0.00061). One of the sites
183 with the lowest sample size in this study (Madagascar, N = 8) was characterized by the highest genetic
184 diversity (h = 0.86 $\pm$ 0.11, π = 0.00214 $\pm$ 0.00166), and the site with the largest sample size (Red Sea, N

185 = 47) was characterized by the lowest genetic diversity (h = 0.20 ± 0.08, π = 0.00031 ± 0.00043), which

186 indicates that differences in genetic diversity are not a result of uneven sampling. The most common

187 haplotype was shared by 267 individuals and detected at every sampling site.

188 The mean number of alleles per microsatellite locus was 18 (range: 13 to 26 alleles), allelic

189 richness was 4.466 (range: 2.764 to 10.112), and observed heterozygosity ranged from 0.285 (Lun 3) to

190 0.883 (B11) (Table 1). Few loci deviated from HWE based on within-site comparisons (8 of 72, $P <$

191 0.02), and no loci were consistently out of equilibrium. No LD was detected based on 180 within-site

192 comparisons after correcting for multiple tests. MICROCHECKER analysis revealed no evidence for

193 scoring error due to stuttering or large allelic dropout. Evidence of null alleles was detected in only 5 of

194 72 comparisons (D2 at central Red Sea, KSA; Christmas Island; Cocos-Keeling Islands; Diego Garcia;

195 French Polynesia). We ran all subsequent analyses excluding or including this locus to mitigate bias; our

196 findings were no different between datasets, so we retained all six microsatellite loci. Overall, there was

197 no consistent evidence for departure from HWE, LD or null alleles across all sampled locations,

198 supporting the decision to retain the entire data set.

199

200 **Population structure analysis**

201 Analyses of molecular variance revealed low but significant population structure for *C. auriga* (mtDNA:

202 Φ_{ST} = 0.027, P < 0.001; microsatellites: F_{ST} = 0.023, P < 0.001 or Jost's D = 0.084). Population pairwise

203 tests revealed that mtDNA haplotype frequencies were significantly different in 37 of 91 comparisons at

204 α = 0.05, but only 12 of 91 comparisons at a corrected α = 0.010 (Appendix S1). Microsatellite

205 frequencies were significantly different in 37 of 66 comparisons at α = 0.05, but only 24 of 66

206 comparisons at a corrected α = 0.010 (Appendix S1). This pattern was driven by the differentiation of

207 Red Sea and Hawaiian samples (Appendix S1), which included most of the significant comparisons at α

208 = 0.05 (58% and 51% at mtDNA and microsatellites, respectively) and an even higher proportion at α =
209 0.010 (100% and 71% at mtDNA and microsatellites, respectively).

210 There were significant but inconsistent patterns of population differentiation elsewhere in the
211 range of *C. auriga*. Diego Garcia is significantly isolated from the eastern Indian Ocean (Christmas and
212 Cocos-Keeling Islands) but not from all the sites in the WIO (Socotra, Madagascar and the Seychelles)
213 in one or both genetic assays (Appendix S1). The two largest samples in the Indian Ocean (Seychelles
214 and Diego Garcia) are significantly isolated from most sites in the Indo-Polynesian Province, but not
215 from the equatorial Caroline Islands and French Polynesia in the southern hemisphere. The strongest
216 inconsistency between genetic assays was observed with the Phoenix Islands sample, which was
217 significantly different in most microsatellite comparisons but not in the mtDNA comparisons (Appendix
218 S1). We detected weak but significant IBD for the full mtDNA datatset (r = 0.178, p = 0.042), but not
219 for the reduced mtDNA datasets where Hawai‘i (r = 0.167, p = 0.082) or Red Sea and Socotra (r =
220 0.129, p = 0.146) were removed from the analysis. IBD for microsatellites was not significant (full
221 dataset: r = 0.023, p = 0.352; Hawai‘i removed: r = 0.268, p = 0.028; Red Sea removed: r = -0.040, p =
222 0.540).

223 STRUCTURE indicated mean probabilities as being highest for *C. auriga* at K = 1, and
224 STRUCTURE HARVESTER identified mean probabilities as being highest at K = 2 (Appendix S2 and
225 S3).Given that the Evanno method is not capable of performing the comparison of K= 1 versus greater
226 values, we accept K = 1 as the most likely value of K. As noted by Evanno *et al.* (2005), STRUCTURE
227 may miss subtle but significant population separations. DAPC analysis confirmed a lack of partitioning
228 between populations, with the exception of the Red Sea and Hawai‘i, which occupied a broader
229 parameter space (*i.e.,* confidence ellipses) and modest overlap with all other sites (Fig. 2).

230

231 **Historical demography**

232 Negative and significant Fu's F_S values were detected in 10 of the 14 sites considered for cyt *b* (Fu's F_S
233 = -7.81 to -1.02; Table 2). The estimates of τ resulted in the Red Sea and Socotra being much older
234 (111,940 to 223,881 years) than all other sites (range: 16,343 to 120,703 years; Table 2). Statistical
235 parsimony networks are consistent with a scenario of low mtDNA differentiation among sites (Fig. 3)
236 and a shallow population history with recent expansion.

237

238 **DISCUSSION**

239 The threadfin butterflyfish has an exceptionally broad distribution that coincides with minimal
240 divergence among sampling sites. Only two peripheral locations, the Red Sea (mtDNA: $\Phi_{ST} = 0.026$, P
241 $= 0.005$; microsatellites: $F_{ST} = 0.010$, $P < 0.045$) and Hawai'i (mtDNA: $\Phi_{ST} = 0.072$, $P = 0.007$;
242 microsatellites: $F_{ST} = 0.133$, $P < 0.001$), were consistently differentiated. Samples from the centre of the
243 range revealed inconsistent population structure across the vast Indo-Polynesian Province. The region
244 from French Polynesia to Western Australia (>200 m depth) has no oceanic gap greater than 800 km,
245 and this almost certainly contributes to genetic cohesiveness (Schultz *et al.,* 2008). Moreover, genetic
246 surveys of dispersive reef organisms are consistent with the boundaries of the Indo-Polynesian Province
247 (Briggs & Bowen, 2012 and references therein; but see Kulbicki *et al.,* 2013). A factor that is frequently
248 invoked to explain high connectivity is the PLD, which is relatively long in butterflyfishes (~40-53 days
249 in this case). Several recent reviews have evaluated the effect of PLD on population genetic structure,
250 yielding a correlation of $r^2 = 0.30$ for a broad spectrum of marine organisms (Selkoe & Toonen, 2011),
251 and $r^2 = 0.22$ for reef fishes (Selkoe *et al.,* 2014). Genetic parentage analysis of *Chaetodon vagabundus*
252 (PLD = 29 to 48 days; Berumen *et al.,* 2012) also found concordance between the level of local retention

253 and PLD. We therefore conclude that the long PLD of *C. auriga* is a factor contributing to minimal
254 divergence, albeit nested within a suite of other physical and biotic factors.
255 There are numerous instances in *C. auriga* where pairwise comparisons between geographical
256 locations are significant in one genetic assay but not the other, most notably with the Phoenix Island
257 sample (Appendix S1). Part of this discrepancy can be attributed to inheritance dynamics of different
258 markers; each may be more sensitive to restrictions on gene flow depending on a variety of demographic
259 conditions (Karl *et al*., 2012). Similar concerns were raised on the relationship between signal and
260 diversity (Jost, 2008; Faubry & Barber, 2012), which differs between nuclear and mitochondrial
261 markers. Other discrepancies can be attributed to the significance level of $P = 0.05$ based on traditional
262 standards and the Narum (2006) correction. In several pairwise comparisons, one genetic assay is just
263 below the significance level, and the other is just above. For this reason we have interpreted pairwise
264 comparisons as significant if one or both assays meet this criterion.
265
266 **Patterns of genetic differentiation**
267 The Red Sea and Socotra in the adjacent Gulf of Aden are significantly divergent in 12 of 13 mtDNA
268 and 9 of 11 microsatellite population comparisons (Appendix S1). This finding is consistent with several
269 recent surveys that show isolation of the Red Sea populations in broadly distributed reef fishes
270 (DiBattista *et al*., 2013). The population-level isolation of the Red Sea is matched by an endemism level
271 of 12.9% in fishes (DiBattista *et al.*, in review A). These partitions are likely promoted by the isolation
272 of the Red Sea during Pleistocene glaciations. The only connection with the Indian Ocean at the Strait of
273 Bab al Mandab is relatively shallow (137 m) and influenced by sea level drops of up to 140 m during
274 glaciations (Rohling *et al*., 2014). Additional oceanographic factors that may isolate the Red Sea include

275 elevated temperature and salinity (Siddall *et al.*, 2004), and cold-water upwelling outside the Red Sea
276 (Kemp, 1998).

277 Population expansion analyses indicates that all *C. auriga* share a common ancestor in the last
278 few hundred thousand years, and that the Red Sea and Socotra host the oldest expansion time. This
279 invokes a hypothetical scenario of recent radiation out of the Red Sea or adjacent areas in response to
280 glacial sea level and climate change, which is consistent with a large proportion of Red Sea endemism
281 spreading to the Gulf of Aden (DiBattista *et al.,* in review A). This post-glacial expansion hypothesis is
282 supported by a lack of genetic (or biogeographical) differentiation between the Red Sea and adjacent
283 Gulf of Aden. In light of this documented isolation, the Red Sea remains an intriguing but understudied
284 region with great potential to inform evolutionary processes in the broader Indo-West Pacific (Berumen
285 *et al.*, 2013). Alternatively, the cold and high nutrient water upwelling just west of Socotra might be the
286 main barrier driving this differentiation (DiBattista *et al.*, in review B).

287 Hawai'i is significantly different in 10 of 13 mtDNA and 11 of 11 microsatellite comparisons
288 (Appendix S1). Hawai'i is one of the most isolated archipelagos in the world with the highest level of
289 endemism in the Pacific (~25%; Randall, 2007). The recurrent trend of genetic distinctness in this region
290 can be attributed to three factors: (1) geographical isolation coupled with oceanographic features that
291 enhance this isolation (Kobayashi, 2006), (2) life history characteristics of the reef biota, including
292 dispersal capabilities (Luiz *et al.,* 2012) and (3) adaptation to environmental conditions in Hawai'i (Bird
293 *et al*., 2012).

294 More subtle patterns of isolation were detected in other locations. Two samples in the Indian
295 Ocean (Seychelles and Diego Garcia) were significantly isolated from most locations in the Indo-
296 Polynesian Province in one or both genetic assays. This is likely a product of the episodic closure of the
297 Indo-Pacific Barrier, a partial land bridge that forms between the Indian and Pacific Oceans during low

298 sea level stands associated with glaciations (Gaither & Rocha, 2013). It is notable that highly dispersive
299 species (as inferred from population genetic comparisons) have little or no structure across this barrier
300 (Craig *et al*., 2007; Horne *et al*., 2008; Reece *et al*., 2011). In contrast, less dispersive species show
301 evolutionary genetic partitions (DiBattista *et al*., 2012b; Gaither & Rocha, 2013). The Threadfin
302 Butterflyfish belongs in the first category.

303 The finding of population structure at the endpoints of the range, and a lack of divergence in the
304 middle, is consistent with other genetic surveys of Indo-Pacific reef fishes. Winters *et al.* (2010)
305 observed genetic homogeneity through most of the Indo-Polynesian Province in the parrotfish *Scarus*
306 *psittacus*, but found isolated populations at Hawai‘i, the Marquesas and the Seychelles. Notably, these
307 regions are also isolated biogeographical provinces as defined by the criteria of >10% endemism (Briggs
308 & Bowen, 2012). Gaither *et al.* (2010) observed a similar pattern in the snapper *Lutjanus kasmira*,
309 reporting that the only isolated populations were in peripheral locations of the WIO and the Marquesas
310 Islands. DiBattista *et al.* (2011) reported genetic homogeneity across the Pacific in the surgeonfish
311 *Acanthurus nigroris*, but an ancient genetic partition at Hawai‘i. Szabo *et al*. (2014) surveyed the
312 goatfish *Parupeneus multifasciatus* across the Pacific, and found a cryptic species at the Marquesas. The
313 threadfin butterflyfish at the Marquesas might also be unique, but our sample size is too small to make
314 this determination. Peripheral isolation and speciation is not the only evolutionary pathway observed in
315 the tropical Indo-Pacific (Cowman & Bellwood, 2013; Gaither & Rocha, 2013), however, it seems to be
316 one of the predominant pathways to speciation (Rocha & Bowen, 2008; Drew & Barber, 2009; Bowen *et*
317 *al*., 2013; Hodge *et al.,* 2014). Tests for IBD were weak or inconclusive, invoking the possibility that
318 divergence of Hawaiian and Red Sea populations is based on founder effects, or that the lack of
319 differentiation across the center of the species range is weakening the IBD signal. While one axis of
320 discrimination for the DAPC analysis separates populations from west to east (Fig. 3), and a nearly

321 perpendicular eigenvector differentiates Hawai‘i from other populations, our estimates of population
322 expansion and genetic diversity do not support founder effects for Hawai‘i.
323

324 **Taxonomic distinction**

325 Reef fishes include many cases of taxonomy based on colouration, especially in butterflyfishes and
326 angelfishes (families Chaetodontidae and Pomacanthidae). Over the past 20 years, several taxonomic
327 distinctions based on colouration have been evaluated with mtDNA and nuclear DNA sequence data.
328 The results have been equivocal, with some genetic lineages aligning with colouration (Drew *et al*.,
329 2010), others showing discordance (Gaither *et al*., 2014) and some groups showing both (McMillan &
330 Palumbi, 1999; Rocha, 2004). As noted by DiBattista *et al*. (2012a), when colour-based taxonomy
331 disagrees with genetic partitions, the latter usually aligns with biogeography.

332 The threadfin butterflyfish has a Red Sea colour morph and proposed subspecies (*C. auriga*
333 *auriga*; Allen, 1979). Randall (1998) recommends a return to subspecies designations in reef fish when
334 morphological and genetic differentiation fall below that observed among congeners, or when
335 interbreeding is likely to be successful. We observed low and inconsistent population genetic
336 differentiation between putative subspecies of *C. auriga*: the genetic partition that included *C. auriga*
337 *auriga* also included individuals identified as *C. auriga setifer* from the central Red Sea and Socotra.
338 Therefore, subspecies designation might not be appropriate in this case, and while we agree with
339 Randall (1998) that evolutionary partitions below the species level are valuable, that criterion does not
340 apply here. Some colour variants may be related to strong sexual selection (*e.g.,* egg spots on cichlids;
341 Santos *et al.,* 2014), or predator avoidance, but there is little evidence linking colour variants to
342 ecological differences in butterflyfishes (Kelley *et al.,* 2013). The bright and stark differences in

colouration, while obvious to the human eye, may reflect evolutionarily labile traits. This does not, however, preclude the possibility that this could be a starting point for diversification.

ACKNOWLEDGEMENTS

This research was supported by NSF grant OCE-0929031 to B.W.B., NOAA MOA No. 2005-008/66882 to R. Toonen, HIMB-NWHI NMSP MOA 2005-008/6682 to M.T.C., 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. and by an NSERC postgraduate fellowship to J.D.D. For specimen collections we thank A. Alexander, T. Alpermann, K. Andersen, P. Barber, C. Braun, R. Coleman, G. Concepcion, A. Connell, T. Daly-Engel, J. Drew, J. Earle, J. Eble, K. Flanagan, M. Gaither, B. Greene, M. Iacchei, S. Jones, S. Karl, R. Kosaki, C. Meyer, G. Nanninga, Y. Papastamatiou, D. Pence, M. Priest, J. Puritz, R. Pyle, J. Reece, D. Robertson, P. Saenz-Agudelo, D. Smith, Z. Szabo, T. Sinclair-Taylor, K. Tenggardjaja, B. Walsh, D. Wegner, I. Willliams, J. Zamzow, the crew of the R.V. *Hi'ialakai* and members of the Reef Ecology Lab at KAUST. We thank Sue Taei at Conservation International, Graham Wragg of the RV Bounty Bay, the Government of Kiribati, including Tukabu Teroroko and the Phoenix Island Protected Area. For support in Socotra, we thank the Ministry of Water and Environment of Yemen, staff at the EPA Socotra, and especially Salah Saeed Ahmed, Fouad Naseeb, and Thabet Abdullah Khamis, as well as Ahmed Issa Ali Affrar for handling general logistics. We thank Frédéric Ramahatratra and le Ministère de la Pêche et des Résources Halieutiques for providing samples from Madagascar. For logistic support elsewhere, we thank Eric Mason at Dream Divers, David Pence, Robert Toonen, Serges Planes, Ben Victor, the Hawai‘i Department of Land and Natural Resources, the Coral Reef Research Foundation and the Papahānaumokuākea Marine National Monument, the Nature Conservancy in Palmyra, U.S. Fish and

366 Wildlife Service, members of the ToBo lab, the Administration of the British Indian Ocean Territory
367 and Charles Sheppard and the KAUST Coastal and Marine Resources Core Lab along with Amr Gusti.
368 We also thank P. Saenz-Agudelo and S. Montanari for their assistance with DAPC analysis, and the
369 Center for Genomics, Proteomics and Bioinformatics at the University of Hawai'i (Manoa Campus).
370 Thanks to editor Gustav Paulay and two anonymous reviewers for comments that improved the
371 manuscript. This is contribution no. 1627 from the Hawai'i Institute of Marine Biology and no. 9437
372 from the School of Ocean and Earth Science and Technology.

373

374 **REFERENCES**

375 Allen, G.R. (1979) *Butterfly and Angelfishes of the World*. New York, Wiley, Volume II, 352 pp.
376 Bandelt, H.J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific
377 phylogenies. *Molecular Biology and Evolution*, **16,** 37–48.
378 Bernardi, G., Holbrook, S.J., Schmitt, R.J., Crane, N.L. & DeMartini, E. (2002) Species boundaries,
379 populations and colour morphs in the coral reef three–spot damselfish (*Dascyllus trimaculatus*)
380 species complex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*,
381 **269,** 599–605.
382 Berumen, M.L. (2005) The importance of juveniles in modelling growth: butterflyfish at Lizard Island.
383 *Environmental Biology of Fishes*, **72,** 409–413.
384 Berumen, M., Rochel, E., Almany, G., Thorrold, S., Jones, G., Pratchett, M., Syms, C. & Planes, S.
385 (2009) Isolation and characterization of 15 polymorphic nuclear microsatellite primers for the
386 widespread Indo-Pacific vagabond butterflyfish, *Chaetodon vagabundus. Molecular Ecology*
387 *Resources*, **9,** 1460–1466.

388 Berumen, M.L., Almany, G.R., Planes, S., Jones, G.P., Sanez-Agudelo,P. & Thorrold, S.R. (2012)
389 Persitence of self-recruitment and patterns of larval connectivity in a marine protected area
390 network. *Ecology and Evolution*, **2,** 444–452.
391 Berumen, M.L., Hoey, A.S., Bass, W.H., Bouwmeester, J., Catania, D., Cochran, J.E.M., Khalil, M.T.,
392 Miyake, S., Mughal, M.R., Spaet, J.L.Y. & Saenz-Agudelo, P. (2013) The status of coral reef
393 ecology research in the Red Sea. *Coral Reefs,* **32,** 737–748.
394 Bird, C.E., Karl, S.A., Smouse, P.E. & Toonen, R.J. (2011) Detecting and measuring genetic
395 differentiation. *Phylogeography and Population Genetics in Crustacea*, **19,** 31–55.
396 Bird, C.E., Fernandez-Silva, I., Skillings, D.J. & Toonen, R.J. (2012) Sympatric speciation in the post
397 "Modern Synthesis" era of evolutionary biology. *Evolutionary Biology*, **39,** 158–180.
398 Blum, S.D. (1989) Biogeography of the Chaetodontidae: an analysis of allopatry among closely related
399 species. *Environmental Biology of Fishes*, **25,** 9–31.
400 Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S. & Robertson, D.R. (2001) Phylogeography of the
401 trumpetfishes (*Aulostomus*): Ring species complex on a global scale. *Evolution*, **55,** 1029–1039.
402 Bowen, B.W., Rocha, L.A., Toonen, R.J., Karl, S.A., Craig, M.T., DiBattista, J.D., Eble, J.A., Gaither,
403 M.R., Skillings, D. & Bird, C.E. (2013) Origins of tropical marine biodiversity. *Trends in*
404 *Ecology and Evolution*, **28,** 359–366.
405 Briggs, J.C. & Bowen. B.W. (2012) A realignment of marine biogeographic provinces with particular
406 reference to fish distributions. *Journal of Biogeography*, **39,** 12–30.
407 Chao, A. & Shen, T.-J. (2010) Program SPADE (Species Prediction And Diversity Estimation).
408 Program and User's Guide published at http://chao.stat.nthu.edu.tw.
409 Cowman, P.F. & Bellwood, D.R. (2013) The historical biogeography of coral reef fishes: global patterns
410 of origination and dispersal. *Journal of Biogeography*, **40,** 209–224.

411 Craig, M., Eble, J.A., Bowen, B.W. & Robertson, D.R. (2007) High genetic connectivity across the
412 Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). *Marine Ecology*
413 *Progress Series*, **334,** 245–254.
414 Craig, M.T., Eble, J.A. & Bowen, B.W. (2010) Origins, ages and population histories: comparative
415 phylogeography of endemic Hawaiian butterflyfishes (genus *Chaetodon*). *Journal of*
416 *Biogeography*, **37,** 2125–2136.
417 Crawford, N.G. (2010) SMOGD: software for the measurement of genetic diversity. *Molecular*
418 *Ecology Resources*, **10,** 556–557.
419 DiBattista, J.D., Wilcox, C., Craig, M.T., Rocha, L.A. & Bowen, B.W. (2011) Phylogeography of the
420 Pacific Blueline Surgeonfish *Acanthurus nigroris* reveals a cryptic species in the Hawaiian
421 Archipelago. *Journal of Marine Biology*, **2011**, Article ID 839134.
422 DiBattista, J.D., Waldrop, E., Bowen, B.W., Schultz, J.K., Gaither, M.R., Pyle, R.L. & Rocha, L.A.
423 (2012a) Twisted sister species of Pygmy Angelfishes: Discordance between taxonomy,
424 coloration, and phylogenetics. *Coral Reefs*, **31,** 839–851.
425 DiBattista, J.D., Craig, M.T., Rocha, L.A., Feldheim, K.A. & Bowen, B.W. (2012b) Phylogeography of
426 the Indo-Pacific butterflyfishes, *Chaetodon meyeri* and *Chaetodon ornatissimus*: Sister species
427 reveal divergent evolutionary histories and discordant results from mtDNA and microsatellites.
428 *Journal of Heredity*, **103,** 617–629.
429 DiBattista, J.D., Berumen, M.L., Gaither, M.R., Rocha, L.A., Eble, J.A., Choat, J.H., Craig, M.T.,
430 Skillings, D.J. & Bowen B.W. (2013) After continents divide: Comparative phylogeography of
431 reef fishes from the Red Sea and Indian Ocean. *Journal of Biogeography*, **40,** 1170–1181.
432 DiBattista, J.D., Roberts, M., Bouwmeester, J., Bowen, B.W., Coker, D.F., Lozano-Cortés, D.F., Choat,

433 J.H., Gaither, M.R., Hobbs, J.P., Kahil, M., Kochzius, M., Myers, R.F., Paulay, G., Robitzch, V.,
434 Saenz-Agudelo, P., Salas, E., Sinclair-Taylor, T.H., Toonen, R.J., Westneat, M., Williams, S.,
435 and Berumen, M.L. (in review A) A review of contemporary patterns of endemism for shallow
436 water reef fauna in the Red Sea. *Journal of Biogeography*.

437 DiBattista, J.D., Choat, J.H., Gaither, M.R., Hobbs, J.P., Lozano-Cortés, D.F., Myers, R.F., Paulay, G.,
438 Rocha, L.A., Toonen, R.J., Westneat, M. & Berumen, M.L. (in review B) On the origin of
439 endemic species in the Red Sea.

440 Drew, J., Allen, G. R., Kaufman, L. E. S. & Barber, P. H. (2008) Endemism and regional color and
441 genetic differences in five putatively cosmopolitan reef fishes. *Conservation Biology*, **22,** 965–
442 975.

443 Drew, J. & Barber, P.H. (2009) Sequential cladogenesis of the reef fish *Pomacentrus moluccensis*
444 (Pomacentridae) supports the peripheral origin of marine biodiversity in the Indo-Australian
445 archipelago. *Molecular Phylogenetics and Evolution*, **53,** 336–339.

446 Drew, J.A., Allen, G.R. & Erdmann, M.V (2010) Congruence between mitochondrial genes and color
447 morphs in a coral reef fish: population variability in the Indo-Pacific damselfish *Chrysiptera rex*
448 (Snyder, 1909). *Coral Reefs*, **29,** 439–444.

449 Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T.
450 & Wilson, A. (2009) *Geneious v4.8*, Available from http://www.geneious.com.

451 Dupanloup, I., Schneider, S. & Excoffier, L. (2002) A simulated annealing approach to define the
452 genetic structure of populations. *Molecular Ecology*, **11,** 2571–2581.

453 Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for
454 visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics*
455 *Resources*, **4,** 359–361.

456 Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the
457 software STRUCTURE: a simulation study. *Molecular Ecology*, **14,** 2611–2620.
458 Excoffier, R., Laval, L.G. & Schneider, S. (2005) *Arlequin ver. 3.0*: An integrated software package for
459 population genetics data analysis. *Evolutionary Bioinformatics Online*, **1,** 47–50.
460 Faubry, S. & Barber, P.H. (2012) Theoretical limits to the correlation between pelagic larval duration
461 and population genetic structure. *Molecular Ecology*, **21,** 3419–3432.
462 Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and
463 background selection. *Genetics*, **147,** 915–925.
464 Gaither, M.R. & Rocha, L.A. (2013) Origins of species richness in the Indo-Malay-Philippine
465 biodiversity hotspot: evidence for the centre of overlap hypothesis. *Journal of Biogeography,* **40,**
466 1638–1648.
467 Gaither, M.R., Toonen, R.J., Robertson, D.R., Planes, S & Bowen, B.W. (2010) Genetic evaluation of
468 marine biogeographic barriers: perspectives from two widespread Indo-Pacific snappers
469 (*Lutjanus* spp.). *Journal of Biogeography*, **37,** 133–147.
470 Gaither, M.R., Schultz, J.K., Bellwood, D., Pyle, R.L., DiBattista, J.D., Rocha, L.A. & Bowen, B.W.
471 (2014) Evolution of pygmy angelfishes: recent divergences, introgression, and the usefulness of
472 color in taxonomy. *Molecular Phylogenetics and Evolution*, **74,** 38–47.
473 Goudet, J. (2001) Fstat, a program to estimate and test gene diversities and fixation indices (v. 2.9.3).
474 See http://www.unil.ch/izea/softwares/fstat.html.
475 Grady, J.M. & Quattro, J.M. (1999) Using character concordance to define taxonomic and
476 conservation units. *Conservation Biology*, **13,** 1004–1007.
477 Hedrick, P.W. (2005) A standardized genetic differentiation measure. *Evolution*, **59,** 1633–1638.
478 Hodge, J.R., Herwerden, L. & Bellwood, D.R. (2014) Temporal evolution of coral reef fishes: global

479 patterns and disparity in isolated locations. *Journal of Biogeography*, **41,** 2115–2127.

480 Hoffman, H.I. & Amos, W. (2005) Microsatellite genotyping errors: detection approaches, common

481 sources and consequences for paternal exclusion. *Molecular Ecology*, **14,** 599–612.

482 Horne, J.B., van Herwerden, L., Choat, J.H. & Robertson, D.R. (2008) High population connectivity

483 across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners.

484 *Molecular Phylogenetics and Evolution*, **49,** 629–638.

485 Jensen, J.L., Bohonak, A.J. & Kelley, S.T. (2005) Isolation by distance, web service. BMC Genetics 6:

486 13. v.3.23 http://ibdws.sdsu.edu/

487 Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal components: a new

488 method for the analysis of genetically structured populations. *BMC Genetics*, **11,** 94.

489 Jost, L.O.U. (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17,** 4015–

490 4026.

491 Karl, S.A., Toonen, R.J., Grant, W.S. & Bowen, B.W. (2012) Common misconceptions in molecular

492 ecology: Echos of the modern synthesis. *Molecular Ecology*, **21,** 4171–4189.

493 Kelley, J.L., Fitzpatrick, J.L. & Merilaita, S. (2013) Spots and stripes: ecology and colour pattern

494 evolution in butterflyfishes. *Proceedings of the Royal Society B: Biological Sciences*, **280,**

495 20122730.

496 Kemp, J. (1998) Zoogeography of the coral reef fishes of the Socotra Archipelago. *Journal of*

497 *Biogeography*, **25,** 919–933.

498 Kobayashi, D.R. (2006) Colonization of the Hawaiian Archipelago via Johnston Atoll: a

499 characterization of oceanographic transport corridors for pelagic larvae using computer

500 simulation. *Coral Reefs*, **25,** 407–417.

501 Kulbicki, M., Parravicini, V., Bellwood, D.R., Arias-Gonzàlez, E., Chabanet, P., Floeter, S.R.,

502 Friedlander, A., McPherson, J., Myers, R.E., Vigliola, L. & Mouillot, D. (2013) Global
503 biogeography of reef fishes: a hierarchical quantitative delineation of regions. *PloS one*, 8,
504 e81847.
505 Lawton, R. J., Pratchett, M. S. & Bay, L. K. (2010) Isolation and characterization of 29 microsatellite
506 loci for studies of population connectivity in the butterflyfishes *Chaetodon trifascialis* and
507 *Chaetodon lunulatus*. *Conservation Genetics Resources*, **2,** 209–213.
508 Lawton, R. J., Pratchett, M. S. & Bay, L. K. (2011) Cross-species amplification of 44 microsatellite loci
509 developed for *Chaetodon trifascialis*, *C. lunulatus* and *C. vagabundus* in 22 related butterflyfish
510 species. *Molecular Ecology Resources*, **11,** 323–327.
511 Leis, J.M. (1989) Larval biology of butterflyfishes (Pisces, Chaetodontidae): what do we really know?
512 *Developments in Environmental Biology of Fishes*, **9,** 87–100.
513 Lessios, H.A. (2008) The great American schism: Divergence of marine organisms after the rise of the
514 Central American isthmus. *Annual Review of Ecology, Evolution, and Systematics*, **39,** 63–91.
515 Luiz, O.J., Madin, J.S., Robertson, D.R., Rocha, L.A., Wirtz, P., & Floeter, S.R. (2012) Ecological traits
516 influencing range expansion across large oceanic dispersal barriers: insights from tropical
517 Atlantic reef fishes. *Proceedings of the Royal Society B: Biological Sciences*, **279,** 1033–1040.
518 McMillan, W.O. & Palumbi, S.R. (1995) Concordant evolutionary patterns among Indo-West Pacific
519 butterflyfishes. *Proceedings of the Royal Society of London*, **260,** 229–236.
520 McMillan, W.O., Weigt, L.A. & Palumbi, S.R. (1999) Color pattern evolution, assortative mating, and
521 genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). *Evolution*, 247–260.
522 Meeker, N.D., Hutchinson, S.A., Ho, L. & Trede, N.S. (2007) Method for isolation of PCR-ready
523 genomic DNA from zebrafish tissues. *Biotechniques*, **43,** 610–614.
524 Messmer, V., Jones, G.P., van Herwerden, L. & Munday, P.L. (2005) Genetic and ecological

525 characterisation of colour dimorphism in a coral reef fish. *Environmental Biology of Fishes*, **74,**
526 175–183.
527 Montanari, S.R., van Herwerden, L., Pratchett, M.S., Hobbs, J.P.A. & Fugedi, A. (2012) Reef fish
528 hybridization: lessons learnt from butterflyfishes (genus *Chaetodon*). *Ecology and Evolution,* **2,**
529 310–328.
530 Narum, S.R. (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics.
531 *Conservation Genetics*, **7,** 783–787.
532 Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25,**
533 1253–1256.
534 Posada, D. & Crandall, K.A. (1998) *Modeltest*: Testing the model of DNA substitution. *Bioinformatics*,
535 **14,** 817–818.
536 Pritchard, J., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus
537 genotype data. *Genetics*, **155,** 945–959.
538 Randall, J.E. (1998) Zoogeography of shore fishes of the Indo-Pacific region. *Zoological Studies*, **37,**
539 227–268.
540 Randall, J.E. (2007) *Reef and Shore Fishes of the Hawaiian Islands*. University of Hawaii Press,
541 Honolulu.
542 Raymond, M. & Rousset, F. (1995) GENEPOP (Version 1.2): Population genetics software for exact
543 tests and ecumenicism. *Journal of Heredity,* **86,** 248–249.
544 Reece, J.S., Bowen, B.W. & Larson, A. (2011) Long larval duration in moray eels (Muraenidae) ensures
545 ocean-wide connectivity despite differences in adult niche breadth. *Marine Ecology Progress*
546 *Series*, **437,** 269–277.
547 Rocha, L.A. & Bowen, B.W. (2008) Speciation in coral reef fishes. *Journal of Fish Biology*, **72,** 1101–

548 1121.

549 Rogers, A.R. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise

550 genetic differences. *Molecular Biology and Evolution*, **9,** 552–569.

551 Rohling, E.J., Foster, G.L., Grant, K.M., Mariino, G., Roberts, A.P., Tamisiae, M.E. & Williams, F.

552 (2014) Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature*,

553 **508**, 477–482.

554 Santos, M.E., Braasch, I., Boileau, N., Meyer, B.S., Sauteur, L., Böhne, A., Belting, H., Affolter, M. &

555 Salzburger, W. (2014) The evolution of cichlid fish egg-spots is linked with a cis-regulatory

556 change. *Nature Communications*, **5**.

557 Schultz, J.K., Pyle, R.L., DeMartini, E. & Bowen, B.W. (2007) Genetic connectivity among color

558 morphs and Pacific archipelagos for the flame angelfish, *Centropyge loriculus*. *Marine Biology*,

559 **151,** 167–175.

560 Schultz, J.K., Feldheim, K.A., Gruber, S.H., Ashley, M.V., McGovern, T. M. & Bowen, B.W. (2008)

561 Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*). *Molecular*

562 *Ecology*, **17,** 5336–5348.

563 Selkoe, K.A. & Toonen, R.J. (2011) Marine connectivity: a new look at pelagic larval duration and

564 genetic metrics of dispersal. *Marine Ecology Progress Series*, **436,** 291–305.

565 Selkoe, K.A., Gaggiotti, O., ToBo Lab, Bowen, B.W. & Toonen, R.J. (2014) Emergent patterns of

566 population genetic structure for a coral reef community. *Molecular Ecology*, **23,** 3064–3079.

567 Siddall, M., Smeed, D.A., Hemleben, C., Rohling, E.J., Schmelzer, I. & Peltier, W.R. (2004)

568 Understanding the Red Sea response to sea level. *Earth and Planetary Science Letters*, **225,** 421–

569 434.

570 Song, C.B., Near, T.J. & Page, J.M. (1998) Phylogenetic relations among percid fishes are inferred from

571 mitochondrial cytochrome *b* DNA sequence data. *Molecular Phylogenetics and Evolution*, **10,**
572 343–353.
573 Szabo, Z., Snelgrove, B., Craig, M.T., Rocha, L.A. & Bowen, B.W. (2014) Phylogeography of the
574 Manybar Goatfish, *Parupeneus multifasciatus*, reveals moderate structure between the Central
575 and North Pacific and a cryptic endemic species in the Marquesas. *Bulletin of Marine Science*,
576 **90,** 49–512.
577 Taberlet, P., Meyer, A. & Bouvert, J. (1992) Unusually large mitochondrial variation in populations of
578 the blue tit, *Parus caeruleus*. *Molecular Ecology*, **1,** 27–36.
579 Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region
580 of mitochondrial-DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10,** 512–
581 526.
582 Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004) MICRO-CHECKER:
583 software for identifying and correcting genotyping errors in microsatellite data. *Molecular*
584 *Ecology Notes*, **4,** 535–538.
585 Winters, K.L., van Herwerden, L, Choat, J.H. & Robertson, D.R. (2010) Phylogeography of the Indo-
586 Pacific parrotfish *Scarus psittacus*: isolation generates distinctive peripheral populations in two
587 oceans. *Marine Biology*, **157,** 1679–1691.
588

589 **SUPPORTING INFORMATION**

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

591 **Appendix S1** Population pairwise Φ_{ST}, F_{ST} or Jost's *D* values for *Chaetodon auriga*.

592 **Appendix S2** Ln P[D] and Delta K for *Chaetodon auriga* from STRUCTURE HARVESTER.

593 **Appendix S3** STRUCTURE bar plot for *Chaetodon auriga.*

BIOSKETCH

The authors' interests 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: J.D.D. produced DNA sequences, analysed these data and led the writing. B.W.B. conceived the design of this study, collected tissue samples and contributed to writing. L.A.R., M.T.C., and M.L.B. contributed to study design, collected tissue samples and contributed to writing. E.W. produced microsatellite data, as well as contributed to the analysis and writing.

Editor: Gustav Paulay

614 **Table 1** Summary characteristics (N_A, number of alleles; A, allelic richness; H_O and H_E, observed and expected heterozygosity; H_{WE},
615 Hardy-Weinberg equilibrium) for six microsatellite loci based on 344 *Chaetodon auriga* specimens collected throughout the Indo-
616 Pacific region. These markers were developed by Berumen *et al*. (2009) and Lawton *et al*. (2010) and later validated for use in
617 Chaetodontidae (Lawton *et al.,* 2011).

Locus	**Annealing temperature (°C)**	**Allelic range (bp)**	N_A	A	H_O	H_E	H_{WE}
Lun 3	58	152-242	22	3.401	0.285	0.303	$P = 0.987$
D117	56	231-303	13	2.764	0.362	0.364	$P = 0.999$
D118	56	149-204	14	3.029	0.428	0.422	$P = 0.310$
D2	56	136-208	19	3.803	0.520	0.623	$P < 0.001$
B11	58	146-218	26	10.112	0.883	0.882	$P = 0.019$
D120	62	218-279	15	3.688	0.517	0.551	$P = 0.551$
Average (SEM)			18.167	4.466	0.499	0.524	

618

619

620

621

622

623

624 **Table 2** Sample size and molecular diversity indices for *Chaetodon auriga* based on mitochondrial DNA (cytochrome *b*) sequence
625 data. Time since the last population expansion event was calculated using a range of mutation rates (1 to 2% per Myr, Bowen *et al.*,
626 2001; Lessios, 2008) and a generation time of 3 years (see Materials and Methods).
627

Collection locality	N^b	H_N	Expansion time (yrs)	Haplotype diversity ($h \pm SD$)	Nucleotide diversity ($\pi \pm SD$)	Fu's F_S
Central Red Sea, KSA (RDS)	47	4	111940-223881	0.20 ± 0.08	0.00031 ± 0.00043	**-2.95**[a]
Socotra, Yemen (SOC)	15	3	111940-223881	0.26 ± 0.14	0.00040 ± 0.00052	**-1.55**
Zanzibar (ZAN)	2	n/a	n/a	n/a	n/a	n/a
Durban, South Africa (DUR)	2	n/a	n/a	n/a	n/a	n/a
Madagascar (MAD)	8	5	60351-120703	0.86 ± 0.11	0.00214 ± 0.00166	-1.92
Republic of Seychelles (SEY)	30	7	23394-46791	0.46 ± 0.11	0.00089 ± 0.00081	**-4.83**
Diego Garcia (DIG)	33	10	32948-65896	0.60 ± 0.10	0.00124 ± 0.00101	**-7.81**
Cocos-Keeling Islands, Aus. (COC)	35	8	32052-64104	0.58 ± 0.09	0.00114 ± 0.00096	**-4.90**
Christmas Island, Aus. (XMA)	36	9	35410-70821	0.62 ±0.09	0.00133 ±0.00106	**-5.53**
Republic of Palau (PAU)	28	8	43955-87910	0.62 ± 0.10	0.00147 ± 0.00114	**-4.21**
Pohnpei, Caroline Islands (CAR)	35	8	22463-44925	0.45 ± 0.10	0.00085 ± 0.00079	**-6.38**
Kanton Atoll, Phoenix Islands (PHO)	16	5	33507-67015	0.61 ± 0.13	0.00108 ± 0.00095	**-2.30**
Hawaiian Islands (HAW)	16	4	33782-67564	0.62 ± 0.10	0.00106 ± 0.00094	-1.02
Johnston Atoll (JON)	1	n/a	n/a	n/a	n/a	n/a
Palmyra Atoll, Line Islands (PAL)	26	6	29813-59627	0.55 ± 0.10	0.00117 ± 0.00098	**-2.63**
Christmas Island, Line Islands (KIR)	31	7	38172-76343	0.62 ± 0.10	0.00142 ± 0.00111	-2.89
French Polynesia (FRP)	26	4	16343-32687	0.34 ± 0.11	0.00055 ± 0.00061	-2.04
All samples	387	37	27093-54185	0.51 ± 0.03	0.00104 ± 0.00087	**-32.66**

628 [a]Numbers in bold are significant, $P < 0.02$ (Fu, 1997).

629 [b]Abbreviations are as follows: Aus., Australia; KSA, Kingdom of Saudi Arabia; N, sample size; H_N, number of haplotypes.

630

631

632

633 # FIGURE LEGENDS

634 **Figure 1** Scaled map indicating collection sites and samples sizes for *Chaetodon auriga* in the Indo-
635 Pacific. Note that several sites were sampled in French Polynesia (Moorea, Society Islands [*N* = 19];
636 Fakarava, Tuamotu Archipelago [*N* = 4]; Nuku Hiva, Marquesas Archipelago [*N* = 3]), the Hawaiian
637 Islands (Big Island [*N* = 2]; Maui [*N* = 1]; Oahu [*N* = 8]; Kauai [*N* = 2]; Laysan [N = 1]; Lisianski [*N* =
638 2]), and the Red Sea (Al Lith [*N* = 27]; Thuwal [*N* = 20], Kingdom of Saudi Arabia) but were grouped
639 for analysis owing to genetic homogeneity within each region (see Methods). Site abbreviations are
640 described in Table 2. Inset photos show the Red Sea morph (*bottom*) and the more widespread morph
641 (*top*) of *C. auriga* (photo credit: L.A.R.).

642

643 **Figure 2** Scatterplot of DAPC performed on six microsatellite loci for 12 populations of *Chaetodon*
644 *auriga*. Populations are shown by colours, numbers (1= Central Red Sea; 2 = Seychelles; 3 = Diego
645 Garcia; 4 = Cocos-Keeling; 5 = Christmas Island, Australia; 6 = Palau; 7 = Pohnpei; 8 = Kanton Atoll; 9
646 = Hawaiian Islands; 10 = Palmyra Atoll; 11 = Line Islands; 12 = French Polynesia) and 95% inertia
647 ellipses. Diamond symbols represent individual genotypes and axes show the first two discriminant
648 functions.

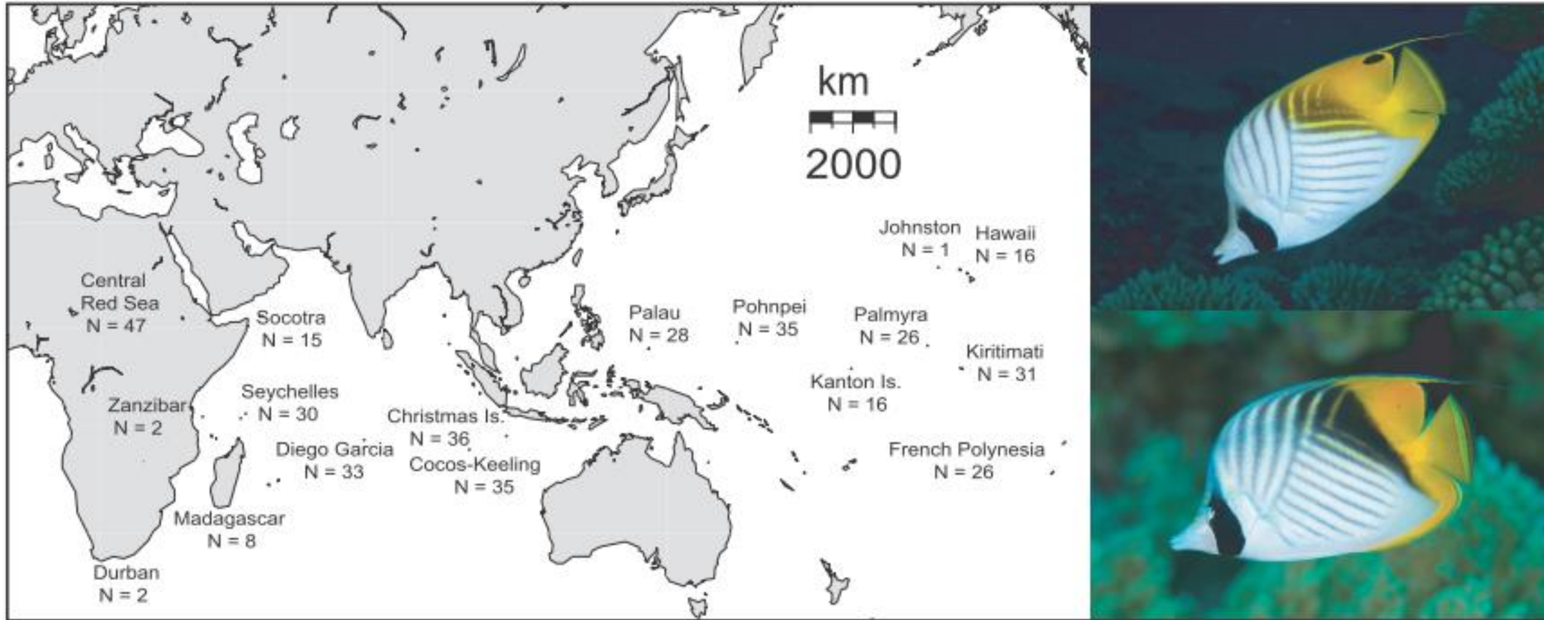
649

650 **Figure 3** Median-joining statistical parsimony networks based on 669 bp of mitochondrial cytochrome *b*
651 sequence data from *Chaetodon auriga* (*N* = 382). Each circle represents a haplotype and its size is
652 proportional to its total frequency. Branches and black crossbars represent a single nucleotide change;
653 colours denote collection location as indicated by the embedded key (as per Fig. 2).

654

655

656 **Figure 1**

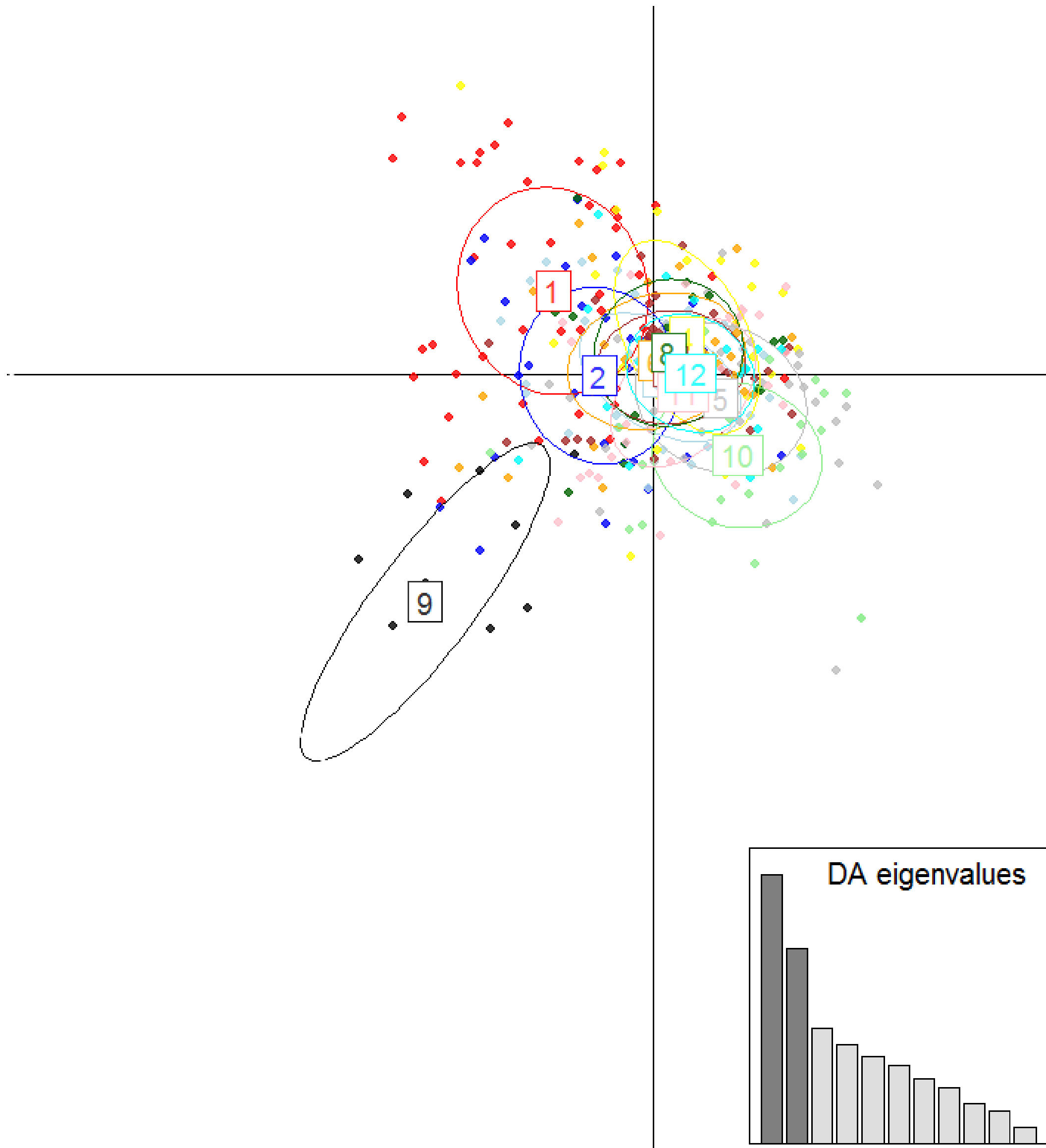


657

658

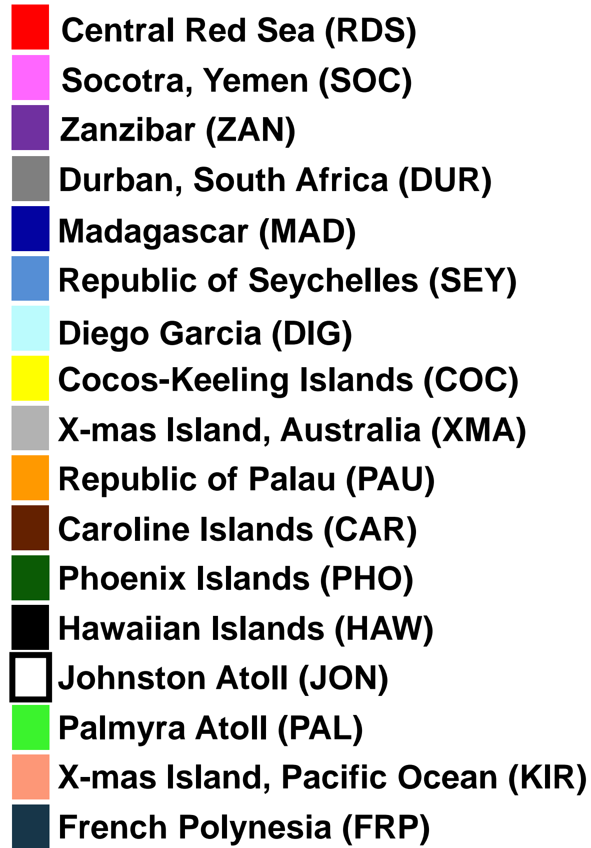
659

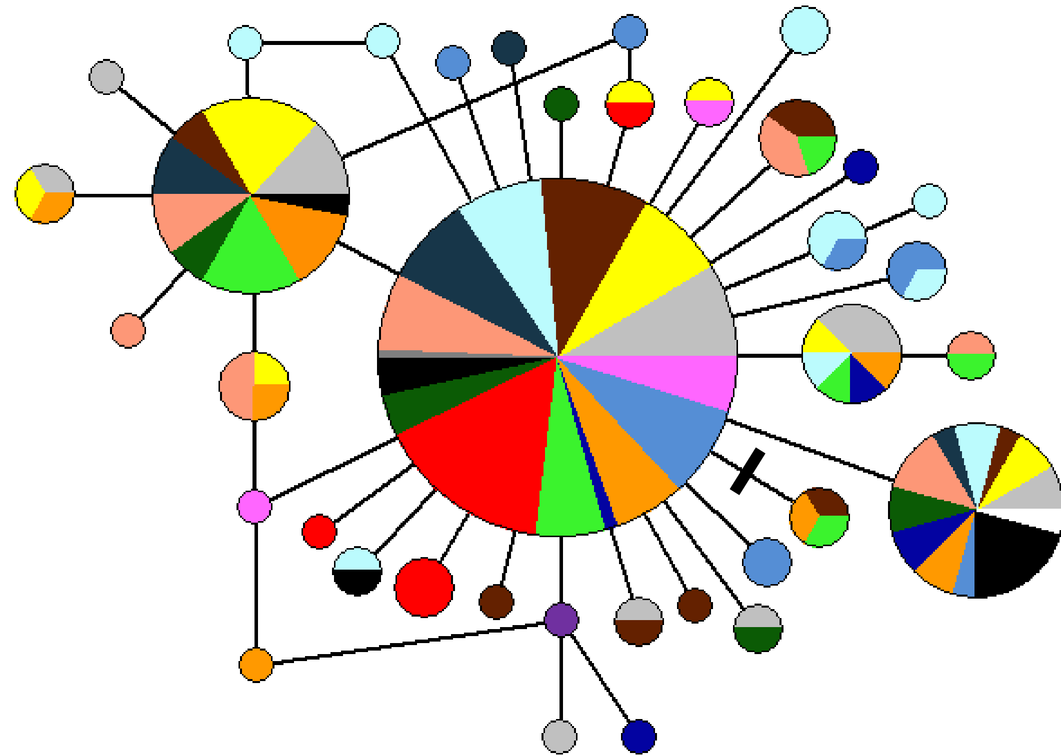
660 **Figure 2**



661

662 **Figure 3**





663

664