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

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>DiBattista, Joseph; Waldrop, Ellen; Rocha, Luiz A.; Craig, Matthew T.; Berumen, Michael L.; Bowen, Brian W.</td>
</tr>
<tr>
<td>Eprint version</td>
<td>Post-print</td>
</tr>
<tr>
<td>DOI</td>
<td>10.1111/jbi.12572</td>
</tr>
<tr>
<td>Publisher</td>
<td>Wiley</td>
</tr>
<tr>
<td>Journal</td>
<td>Journal of Biogeography</td>
</tr>
<tr>
<td>Rights</td>
<td>Archived with thanks to Wiley; This file is an open access version redistributed from: <a href="https://espace.curtin.edu.au/bitstream/20.500.11937/52512/2/251278.pdf">https://espace.curtin.edu.au/bitstream/20.500.11937/52512/2/251278.pdf</a></td>
</tr>
<tr>
<td>Download date</td>
<td>2023-11-01 12:05:16</td>
</tr>
<tr>
<td>Link to Item</td>
<td><a href="http://hdl.handle.net/10754/566080">http://hdl.handle.net/10754/566080</a></td>
</tr>
</tbody>
</table>
Blinded by the bright: A lack of congruence between colour morphs, phylogeography and taxonomy for a cosmopolitan Indo-Pacific butterflyfish, Chaetodon auriga

Joseph D. DiBattista¹²*, Ellen Waldrop³, Luiz A. Rocha⁴, Matthew T. Craig⁵, Michael L. Berumen¹, Brian W. Bowen³

¹Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia, ²Department of Environment and Agriculture, Curtin University, PO Box U1987, Perth, WA 6845, Australia, ³Hawai‘i Institute of Marine Biology, Kāne‘ohe, HI 96744, USA, ⁴Section of Ichthyology, California Academy of Sciences, San Francisco, CA 94118, USA, ⁵Southwest Fisheries Science Center, NOAA National Marine Fisheries Service, La Jolla, CA 92037, USA

*Correspondence: Joseph D. DiBattista, Department of Environment and Agriculture, Curtin University, PO Box U1987, Perth, WA 6845, Australia

E-mail: josephdibattista@gmail.com
ABSTRACT

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

Location Red Sea, Indian Ocean and Pacific Ocean.

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

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

Main conclusions We find little genetic evidence to support an evolutionary partition of a previously
proposed Red Sea subspecies. The oldest estimate of population expansion in the Red Sea and adjacent Gulf of Aden indicates a putative refuge in this region during Pleistocene glacial cycles. The finding of population separations at the limits of the range, in the Red Sea and Hawai‘i, is consistent with peripheral speciation.

Keywords
Coral reef fish, marine biogeography, microsatellite, mitochondrial DNA, population expansion time, subspecies

INTRODUCTION
Colouration plays an important role in the taxonomic classification of reef fishes and is frequently the sole character used to distinguish closely related species. Its evolutionary significance, however, is uncertain (McMillan et al., 1999; Bernardi et al., 2002), since colour variation can be a result of phenotypic plasticity rather than reproductive isolation (Grady & Quattro, 1999). In addition, colouration may evolve faster than morphological and genetic characters (Schultz et al., 2007).

Colour polymorphisms within the same species are relatively common (e.g. brown dottyback, Messmer et al., 2005) and several mechanisms have been proposed to explain their existence. Many reef fishes are distinguished primarily by colour, yet colouration is not necessarily a species-specific diagnostic character, particularly for widespread species (flame angelfish, Schultz et al., 2007; King Demoiselle, Drew et al., 2008). In some cases there is greater concordance between genetics and geography than between genetics and colouration (McMillan & Palumbi, 1995; DiBattista et al., 2012a).

Like many reef fishes, the butterflyfishes (family: Chaetodontidae) have a spectacular variety of colour patterns, and there appears to be a link between species diversification and colour variation
(Blum, 1989). The significance of colour in this group, however, must be interpreted with caution. For example, McMillan et al. (1999) observed colour pattern evolution associated with genetic divergence in *Chaetodon multicinctus*, a Hawaiian endemic, but not in its two sister species (*C. punctatofasciatus* and *C. pelewensis*) distributed across the Indo-West Pacific.

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

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

1) Characterize genetic diversity across the species range to resolve demographic histories, with particular emphasis on the Red Sea;

2) Define population genetic structure across the range to resolve the relationship between genetic partitions and biogeographical provinces.
MATERIALS AND METHODS

Sample collection

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

Mitochondrial DNA sequencing

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

ARLEQUIN 3.5.1.2 (Excoffier *et al.*, 2005) was used to calculate haplotype (h) and nucleotide diversity (π), as well as to test for population structure. Genetic differentiation among sampling sites was first estimated with analysis of molecular variance (AMOVA) based on pairwise comparisons of sample groups; deviations from null distributions were tested with non-parametric permutation procedures (*N* = 99,999). Pairwise *Φ*st statistics were also generated in ARLEQUIN, significance tested by permutation.
(\(N = 99,999\)) and \(P\)-values adjusted according to the modified false discovery rate (FDR) method (Narum, 2006). Patterns of significant genetic differentiation were congruent between FDR and more conservative methods of correction (i.e. Bonferonni; data not shown). Multiple sites within French Polynesia, the Hawaiian Islands and the Red Sea were grouped based on preliminary findings of genetic homogeneity. We used the Isolation by Distance (IBD) Web Service 3.23 to detect correlations between geographic and genetic distances (Jensen et al., 2005). To mitigate any false positives, we tested: 1) the whole range, 2) the range minus Red Sea/Socotra and 3) the range minus Hawai‘i.

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

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

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

For each locus the mean number of alleles ($N_A$), observed ($H_O$) and expected ($H_E$) heterozygosities, Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were assessed with GENEPOP 4.2.2 and ARLEQUIN. Allelic richness was determined with FSTAT 2.9.3.2. Significance levels for multiple comparisons were adjusted using false discovery rate method (Narum, 2006). MICRO-CHECKER 2.2.3 was used to identify genotyping errors including null alleles, allelic dropouts and stutter peaks (Van Oosterhout et al., 2004); significance levels for multiple comparisons were adjusted using the sequential Bonferroni correction with default settings. Overall population structure and pairwise comparisons ($F_{ST}$ calculations) were estimated with ARLEQUIN. To facilitate comparisons to other studies, an additional diversity measure, Jost’s $D$ (Jost, 2008), was estimated using SPADE (Chao et al., 2008). This metric compensates for the downward bias in $F_{ST}$ produced by within-population heterozygosity, a recurring problem with microsatellite markers (Bird et al., 2011). IBD tests were further conducted on the microsatellite dataset as outlined above.
STRUCTURE 2.3.2 was used to assign individuals to genetic clusters (populations) without bias from geographical locations (Pritchard et al., 2000). STRUCTURE uses a Bayesian approach to assign individual multi-locus genotypes to clusters ($K$) by minimising deviations from Hardy-Weinberg and linkage equilibrium. The most likely number of clusters was identified by testing the probability of $K = 1$ to $K = 12$. Analyses were repeated five times and the results averaged. Each run consisted of 1,000,000 MCMC repetitions, a burn-in of 10,000 iterations and correlated allele frequencies and admixed populations were assumed (as per DiBattista et al., 2012b). STRUCTURE HARVESTER 0.6.94 identified the most likely $K$ value (genetic groups) (Evanno et al., 2005; Earl & vonHoldt, 2012).

A discriminant analysis of principal components (DAPC; Jombart et al., 2010) was also run on all loci to investigate the relationship between genotype and geographical location. The number of principal components retained for genotypic variability was equal to the number of individuals divided by three; the number of DA eigenvectors corresponded to the number of populations minus one. Although different from the admixture coefficients of STRUCTURE, DAPC can still be interpreted as proximities of individuals to different clusters based on the retained discriminant functions.

RESULTS

Molecular characteristics

Cyt $b$ sequences from *C. auriga* included 33 haplotypes (4 to 10 within-sites), with haplotype and nucleotide diversity ranging from $h = 0.20$ to 0.86 and $\pi = 0.00031$ to 0.00214 (Table 2). Haplotype and nucleotide diversity was almost twice as high at all other sites compared to the Red Sea ($h = 0.20 \pm 0.08$, $\pi = 0.00031 \pm 0.00043$) and French Polynesia ($h = 0.34 \pm 0.11$, $\pi = 0.00055 \pm 0.00061$). One of the sites with the lowest sample size in this study (Madagascar, $N = 8$) was characterized by the highest genetic diversity ($h = 0.86 \pm 0.11$, $\pi = 0.00214 \pm 0.00166$), and the site with the largest sample size (Red Sea, $N$
= 47) was characterized by the lowest genetic diversity ($h = 0.20 \pm 0.08$, $\pi = 0.00031 \pm 0.00043$), which indicates that differences in genetic diversity are not a result of uneven sampling. The most common haplotype was shared by 267 individuals and detected at every sampling site.

The mean number of alleles per microsatellite locus was 18 (range: 13 to 26 alleles), allelic richness was 4.466 (range: 2.764 to 10.112), and observed heterozygosity ranged from 0.285 (Lun 3) to 0.883 (B11) (Table 1). Few loci deviated from HWE based on within-site comparisons (8 of 72, $P < 0.02$), and no loci were consistently out of equilibrium. No LD was detected based on 180 within-site comparisons after correcting for multiple tests. MICROCHECKER analysis revealed no evidence for scoring error due to stuttering or large allelic dropout. Evidence of null alleles was detected in only 5 of 72 comparisons (D2 at central Red Sea, KSA; Christmas Island; Cocos-Keeling Islands; Diego Garcia; French Polynesia). We ran all subsequent analyses excluding or including this locus to mitigate bias; our findings were no different between datasets, so we retained all six microsatellite loci. Overall, there was no consistent evidence for departure from HWE, LD or null alleles across all sampled locations, supporting the decision to retain the entire data set.

**Population structure analysis**

Analyses of molecular variance revealed low but significant population structure for *C. auriga* (mtDNA: $\Phi_{ST} = 0.027$, $P < 0.001$; microsatellites: $F_{ST} = 0.023$, $P < 0.001$ or Jost’s $D = 0.084$). Population pairwise tests revealed that mtDNA haplotype frequencies were significantly different in 37 of 91 comparisons at $\alpha = 0.05$, but only 12 of 91 comparisons at a corrected $\alpha = 0.010$ (Appendix S1). Microsatellite frequencies were significantly different in 37 of 66 comparisons at $\alpha = 0.05$, but only 24 of 66 comparisons at a corrected $\alpha = 0.010$ (Appendix S1). This pattern was driven by the differentiation of Red Sea and Hawaiian samples ( Appendix S1), which included most of the significant comparisons at $\alpha$.
= 0.05 (58% and 51% at mtDNA and microsatellites, respectively) and an even higher proportion at $\alpha = 0.010$ (100% and 71% at mtDNA and microsatellites, respectively).

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

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

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

DISCUSSION

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

There are numerous instances in *C. auriga* where pairwise comparisons between geographical locations are significant in one genetic assay but not the other, most notably with the Phoenix Island sample (Appendix S1). Part of this discrepancy can be attributed to inheritance dynamics of different markers; each may be more sensitive to restrictions on gene flow depending on a variety of demographic conditions (Karl *et al*., 2012). Similar concerns were raised on the relationship between signal and diversity (Jost, 2008; Faubry & Barber, 2012), which differs between nuclear and mitochondrial markers. Other discrepancies can be attributed to the significance level of *P* = 0.05 based on traditional standards and the Narum (2006) correction. In several pairwise comparisons, one genetic assay is just below the significance level, and the other is just above. For this reason we have interpreted pairwise comparisons as significant if one or both assays meet this criterion.

Patterns of genetic differentiation

The Red Sea and Socotra in the adjacent Gulf of Aden are significantly divergent in 12 of 13 mtDNA and 9 of 11 microsatellite population comparisons (Appendix S1). This finding is consistent with several recent surveys that show isolation of the Red Sea populations in broadly distributed reef fishes (DiBattista *et al*., 2013). The population-level isolation of the Red Sea is matched by an endemism level of 12.9% in fishes (DiBattista *et al*., in review A). These partitions are likely promoted by the isolation of the Red Sea during Pleistocene glaciations. The only connection with the Indian Ocean at the Strait of Bab al Mandab is relatively shallow (137 m) and influenced by sea level drops of up to 140 m during glaciations (Rohling *et al*., 2014). Additional oceanographic factors that may isolate the Red Sea include
elevated temperature and salinity (Siddall et al., 2004), and cold-water upwelling outside the Red Sea (Kemp, 1998).

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

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

More subtle patterns of isolation were detected in other locations. Two samples in the Indian Ocean (Seychelles and Diego Garcia) were significantly isolated from most locations in the Indo-Pacific Province in one or both genetic assays. This is likely a product of the episodic closure of the Indo-Pacific Barrier, a partial land bridge that forms between the Indian and Pacific Oceans during low
sea level stands associated with glaciations (Gaither & Rocha, 2013). It is notable that highly dispersive species (as inferred from population genetic comparisons) have little or no structure across this barrier (Craig et al., 2007; Horne et al., 2008; Reece et al., 2011). In contrast, less dispersive species show evolutionary genetic partitions (DiBattista et al., 2012b; Gaither & Rocha, 2013). The Threadfin Butterflyfish belongs in the first category.

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

### Taxonomic distinction

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

The threadfin butterflyfish has a Red Sea colour morph and proposed subspecies (*C. auriga auriga*; Allen, 1979). Randall (1998) recommends a return to subspecies designations in reef fish when morphological and genetic differentiation fall below that observed among congeners, or when interbreeding is likely to be successful. We observed low and inconsistent population genetic differentiation between putative subspecies of *C. auriga*: the genetic partition that included *C. auriga auriga* also included individuals identified as *C. auriga setifer* from the central Red Sea and Socotra. Therefore, subspecies designation might not be appropriate in this case, and while we agree with Randall (1998) that evolutionary partitions below the species level are valuable, that criterion does not apply here. Some colour variants may be related to strong sexual selection (*e.g.*, egg spots on cichlids; Santos et al., 2014), or predator avoidance, but there is little evidence linking colour variants to 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, HIMP-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
Wildlife Service, members of the ToBo lab, the Administration of the British Indian Ocean Territory and Charles Sheppard and the KAUST Coastal and Marine Resources Core Lab along with Amr Gusti.

We also thank P. Saenz-Agudelo and S. Montanari for their assistance with DAPC analysis, and the Center for Genomics, Proteomics and Bioinformatics at the University of Hawai‘i (Manoa Campus).

Thanks to editor Gustav Paulay and two anonymous reviewers for comments that improved the manuscript. This is contribution no. 1627 from the Hawai‘i Institute of Marine Biology and no. 9437 from the School of Ocean and Earth Science and Technology.

REFERENCES


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

endemic species in the Red Sea.

975.

Drew, J. & Barber, P.H. (2009) Sequential cladogenesis of the reef fish *Pomacentrus moluccensis*
(Pomacentridae) supports the peripheral origin of marine biodiversity in the Indo-Australian

morphs in a coral reef fish: population variability in the Indo-Pacific damselfish *Chrysiptera rex*


Dupanloup, I., Schneider, S. & Excoffier, L. (2002) A simulated annealing approach to define the

visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics*
*Resources*, 4, 359–361.


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


**SUPPORTING INFORMATION**

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

**Appendix S1** Population pairwise $\Phi_{ST}$, $F_{ST}$ or Jost’s $D$ values for *Chaetodon auriga*.

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

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

<table>
<thead>
<tr>
<th>Locus</th>
<th>Annealing temperature (°C)</th>
<th>Allelic range (bp)</th>
<th>$N_A$</th>
<th>$A$</th>
<th>$H_O$</th>
<th>$H_E$</th>
<th>$H_{WE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Lun 3</em></td>
<td>58</td>
<td>152-242</td>
<td>22</td>
<td>3.401</td>
<td>0.285</td>
<td>0.303</td>
<td>$P = 0.987$</td>
</tr>
<tr>
<td><em>D117</em></td>
<td>56</td>
<td>231-303</td>
<td>13</td>
<td>2.764</td>
<td>0.362</td>
<td>0.364</td>
<td>$P = 0.999$</td>
</tr>
<tr>
<td><em>D118</em></td>
<td>56</td>
<td>149-204</td>
<td>14</td>
<td>3.029</td>
<td>0.428</td>
<td>0.422</td>
<td>$P = 0.310$</td>
</tr>
<tr>
<td><em>D2</em></td>
<td>56</td>
<td>136-208</td>
<td>19</td>
<td>3.803</td>
<td>0.520</td>
<td>0.623</td>
<td>$P &lt; 0.001$</td>
</tr>
<tr>
<td><em>B11</em></td>
<td>58</td>
<td>146-218</td>
<td>26</td>
<td>10.112</td>
<td>0.883</td>
<td>0.882</td>
<td>$P = 0.019$</td>
</tr>
<tr>
<td><em>D120</em></td>
<td>62</td>
<td>218-279</td>
<td>15</td>
<td>3.688</td>
<td>0.517</td>
<td>0.551</td>
<td>$P = 0.551$</td>
</tr>
<tr>
<td><strong>Average</strong> (SEM)</td>
<td></td>
<td></td>
<td></td>
<td>18.167</td>
<td>4.466</td>
<td>0.499</td>
<td>0.524</td>
</tr>
</tbody>
</table>
Table 2 Sample size and molecular diversity indices for *Chaetodon auriga* based on mitochondrial DNA (cytochrome *b*) sequence data. Time since the last population expansion event was calculated using a range of mutation rates (1 to 2% per Myr, Bowen *et al.*, 2001; Lessios, 2008) and a generation time of 3 years (see Materials and Methods).

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

*a*Numbers in bold are significant, *P* < 0.02 (Fu, 1997).

*b*Abbreviations are as follows: Aus., Australia; KSA, Kingdom of Saudi Arabia; *N*, sample size; *HN*, number of haplotypes.
**FIGURE LEGENDS**

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

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

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

- Central Red Sea (RDS)
- Socotra, Yemen (SOC)
- Zanzibar (ZAN)
- Durban, South Africa (DUR)
- Madagascar (MAD)
- Republic of Seychelles (SEY)
- Diego Garcia (DIG)
- Cocos-Keeling Islands (COC)
- X-mas Island, Australia (XMA)
- Republic of Palau (PAU)
- Caroline Islands (CAR)
- Phoenix Islands (PHO)
- Hawaiian Islands (HAW)
- Johnston Atoll (JON)
- Palmyra Atoll (PAL)
- X-mas Island, Pacific Ocean (KIR)
- French Polynesia (FRP)