Phylogeography of Two Closely Related Indo-Pacific Butterflyfishes Reveals Divergent Evolutionary Histories and Discordant Results from mtDNA and Microsatellites

JOSEPH D. DI BATTISTA, LUIZ A. ROCHA, MATTHEW T. CRAIG, KEVIN A. FELDHEIM, AND BRIAN W. BOWEN

From the Hawai‘i Institute of Marine Biology, Kāne‘ohe, HI (Joseph D. DiBattista); the Section of Ichthyology, California Academy of Sciences, San Francisco, CA (Luiz A. Rocha); the Department of Marine Sciences and Environmental Studies, University of San Diego, San Diego, CA (Matthew T. Craig); the Field Museum of Natural History, Pritzker Laboratory for Molecular Systematics and Evolution, Chicago, IL (Kevin A. Feldheim); and the Hawai‘i Institute of Marine Biology, Kāne‘ohe, HI (Brian W. Bowen).

Address correspondence to Joseph D. DiBattista at the address above, or e-mail: joseph99@hawaii.edu.

Abstract

Marine biogeographic barriers can have unpredictable consequences, even among closely related species. To resolve phylogeographic patterns for Indo-Pacific reef fauna, we conducted range-wide surveys of sister species, the scrawled butterflyfish (*Chaetodon meyeri*, *N* = 134) and the ornate butterflyfish (*Chaetodon ornatissimus*, *N* = 296), using mitochondrial DNA cytochrome *b* sequences and 10 microsatellite loci. The former is distributed primarily in the Indian Ocean but also extends to the Line Islands in the Central Pacific, whereas the latter is distributed primarily in the Central-West Pacific (including Hawaii and French Polynesia) but extends to the eastern margin of the Indian Ocean. Analyses of molecular variance and Bayesian STRUCTURE results revealed 1 range-wide group for *C. meyeri* and 3 groups for *C. ornatissimus*: 1) eastern Indian Ocean and western Pacific, 2) Central Pacific, and 3) Hawaii. Estimates of the last population expansion were much more recent for *C. meyeri* (61,500 to 95,000 years) versus *C. ornatissimus* (184,700 to 286,300 years). Despite similarities in ecology, morphology, life history, and a broadly overlapping distribution, these sister species have divergent patterns of dispersal and corresponding evolutionary history. The mtDNA and microsatellite markers did not provide concordant results within 1 of our study species (*C. meyeri*), or in 7 out of 12 other cases of marine fishes in the published literature. This discordance renews caution in relying on one or a few markers for reconstructing historical demography.

Key words: biogeography, *Chaetodon meyeri*, *Chaetodon ornatissimus*, coral reef fish, gene flow, marine connectivity

Physical marine barriers can isolate populations and provide the starting point for speciation, as evidenced by phylogeographic studies of sister taxa divided by the Isthmus of Panama over 3 million years ago (e.g., Lessios 2008; also see Jordan 1908). Another notable physical barrier (albeit not complete) is the episodic separation between the tropical Indian and Pacific Oceans at the Sunda Shelf owing to geological-climatic events (Indo-Pacific Barrier, IPB; Briggs 1974; Rocha et al. 2007). Repeated lowering of sea level by up to a 120 m during the Pleistocene epoch exposed coral reef, altered the direction of sea surface currents, and limited inter-ocean dispersal across this region (Porter 1989). The IPB has therefore been invoked to explain sister species pairs distributed in the Indian and Pacific Oceans with concordant geographic and genetic boundaries (McMillan and Palumbi 1995; Benzie 1999). This barrier is also thought to be responsible for within-species genetic divisions observed in taxa distributed across the Indian and Pacific Ocean (Craig et al. 2007; Gaiter et al. 2010; but see Horne et al. 2008).

Physical barriers to dispersal also include distance over deep water, which clearly has the potential to limit exchange between populations and incipient species. Indeed, rare dispersal events to oceanic archipelagos, followed by isolation over long periods of time, are important evolutionary
processes (Paulay and Meyer 2002; Drew and Barber 2009; Malay and Paulay 2010). Isolated archipelagos are also less prone to hybridization between resident and migrant fish given the rarity of such dispersal events. As an example, the proportion of endemic shore fishes in the Hawaiian archipelago is one of the highest in the oceans (25%, Randall 2007), which is attributed to the remoteness of this island chain.

Marine barriers based solely on distance, however, do not always translate into reduced gene flow. For example, even the widest marine biogeographic barrier (Eastern Pacific Barrier, EPB) is sporadically permeable to dispersing fish propagules (Robertson et al. 2004). A survey of 20 shore fishes across the 5000 km EPB (separating the Eastern from the Central Pacific) found that only 20% were divergent by population genetic criteria (Lessios and Robertson 2006). Sustained, or even episodic enhancement of oceanographic currents (i.e., El-Niño or La-Niña events, Robertson et al. 2004) may permit dispersal and colonization in the face of such obstacles. Life-history traits for some reef fishes, such as extended larval stages (Reece et al. 2010), strong swimming ability (Selkoe et al. 2008), or specific reproductive strategies (i.e., r-selected species, White et al. 2009) can also enhance intraspecific gene flow.

Correlations between habitat specificity and genetic connectivity indicate that speciation in the sea does not require physical barriers (Rocha et al. 2005; Taylor and Hellberg 2005; Crow et al. 2010; Bird et al. 2011). Species diversification may be driven in part by selection for divergent ecological conditions in the face of ongoing gene flow (i.e., ecological speciation, Weissing et al. 2011). As one example, molecular surveys of a diverse reef fish family (Labridae) found that genetic differentiation was closely tracked by environmental differences throughout the western Atlantic, but not necessarily consistent with existing barriers to dispersal (Rocha et al. 2005). Additional phylogenies of reef fish provide examples of sister taxa with overlapping distributions (gobies, Dawson et al. 2002; seahorses, Jones et al. 2003; grunts, Rocha et al. 2008; rockfish, Crow et al. 2010), and yet such species boundaries are maintained despite rampant hybridization in some cases (Crow et al. 2007).

To elucidate phylogeographic patterns in the oceans, we focus on two putative sister species (Bellwood et al. 2010), the scrawled butterflyfish (Chaetodon meyeri Bloch and Schneider 1801) and the ornate butterflyfish (Chaetodon ornatissimus Cuvier 1831). C. meyeri occurs across the Indian Ocean to the Central Pacific (into the Line Islands but not Hawaii), whereas C. ornatissimus occurs from the eastern Indian Ocean to the Central Pacific, including Hawaii and French Polynesia (Figure 1). Hence these species collectively traverse 3 marine biogeographic provinces: the western Indian Ocean, the vast Indo-Polynesian Province that extends from the central Indian Ocean to Polynesia, and the isolated Hawaiian Province (Briggs 1974; Briggs and Bowen 2012). Both species are additionally distributed across the transitional region between the Indian and Pacific Ocean, and thus potentially informative with regards to the influence of the IPB on genetic structure. These closely related taxa are obligate coralivores and therefore dependent on healthy, shallow-water reef habitat (Myers 1991). These 2 members of the family Chaetodontidae also have similar morphology (overlapping gill raker, scale, and fin-ray counts), life-history characteristics, habitat preferences, (possibly) dispersal potential, and are known to hybridize in parts of their range (e.g., Christmas Island, Hobbs et al. 2009). Hybrid individuals of intermediate color pattern, however, are exceedingly rare outside of the eastern Indian Ocean and parts of Indonesia (R.L. Pyle, pers. comm.).

The 2 species considered here provide an excellent system to explore the influence of historical biogeographic barriers in the Indo-Pacific region. Our study was therefore designed to address 3 primary questions. First, is there evidence of within-species genetic partitions across the Indo-Polynesian
Province for each of these butterflyfish? Second, if so, what is the magnitude of such genetic partitions across the IPB? Third, is the genetic structure of these butterflyfish consistent with population contraction and expansion events, and if so, what is their relative timing (i.e., recent versus ancient)? Given the ecological similarity of these butterflyfish, concordant phylogeographic patterns are expected if these species were subjected to similar vicariant events. Patterns of genetic isolation should also be similar between mtDNA and microsatellite markers in reef fishes, which lack the confounding influence of sex-biased dispersal (Pardini et al. 2001, Bowen et al. 2005; but see Miller-Sims et al. 2008), have large effective population sizes (N_e; for review see Hare et al. 2011), and chromosomal sex determination (Devlin and Nagahama 2002). Discordance between results from mtDNA and microsatellites may require researchers to reevaluate the inheritance dynamics of each class of markers, especially as it pertains to the power to detect population structure (Karl et al. 2012).

**Materials and Methods**

**Sample Collection**

A total of 134 and 296 tissue samples of *Chaetodon meyeri* and *Chaetodon ornatus* respectively, were collected at 10 sites with pole spears while scuba diving or snorkeling between 2006 and 2010 (Figure 1, Table 1). Although Hawaiian collections include multiple locations within the 2500 km archipelago, these were combined for analysis owing to low sample sizes at most individual sites and no genetic differentiation (Kauai, N = 11; French Frigate Shoals, N = 2; Ni‘ihau, N = 1; Oahu, N = 6; Big Island, N = 41). Collectors were aware of the possibility of hybrids between these 2 species in the eastern Indian Ocean, and coloration traits were used to assure that no hybrids were included in this study. Collected tissue was preserved in a saturated salt-DMSO solution, total genomic DNA extracted using a “HotSHOT” protocol (Meeker et al. 2007), and samples subsequently stored at −20 °C.

**Mitochondrial DNA Analysis**

A731 basepair (bp) segment of mtDNA cytchrome b/ (cyt b) gene was amplified using heavy-strand (5’ - GTGACTTGAAGATGGAAGTATCATTCGGGTATTTGATG - 3’, Song et al. 1998) and light-strand primers (5' - AATAGGAATATCATTCGGGTATTTGATG - 3’, Taberlet et al. 1992). Polymerase chain reaction (PCR) mixes contained BioMix (BioMix Red; Bioline Ltd., London, UK), 0.26 μM of each primer, and 5 to 50 ng template DNA in 15 μl total volume. Thermal cycling PCR used the following parameters: initial denaturing step at 95 °C for 3 min, then 35 cycles of amplification (30 s of denaturing at 94 °C, 45 seconds of annealing at 50 °C, and 45 seconds of extension at 72 °C), followed by a final extension at 72 °C for 10 min.

PCR products were visualized through 1.5% agarose gel electrophoresis and purified by incubating with 0.75 units of exonuclease I and 0.5 units of shrimp alkaline phosphatase (ExoSAP; USB, Cleveland, OH, USA) per 7.5 μl PCR product at 37 °C for 60 min, followed by 85 °C for 15 min. All samples were then sequenced in the forward direction (and reverse direction for questionable haplotypes, N = 4) with fluorescently labeled dye terminators (BigDye version 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 version 4.8.4 (Drummond et al. 2009); unique mtDNA cyt b haplotypes were deposited in GenBank (accession numbers: HQ329510 to HQ329585).

ARLEQUIN version 3.1 software (Excoffier et al. 2005) was used to calculate haplotype (b) and nucleotide diversity (π), as well as to test for range-wide patterns of population structure for each species separately. Genetic differentiation among sampling sites was first estimated with an analysis of molecular variance (AMOVA, Excoffier et al. 1992) without group subdivisions; deviations from null distributions were tested with nonparametric permutation procedures (N = 99 999). Population pairwise ΦST statistics were also generated in ARLEQUIN; significance was tested by permutation (N = 99 999) and P-values adjusted according to the modified false discovery rate method (as per Narum 2006). A Mantel test was performed in ARLEQUIN to detect positive correlations between geographic and genetic distance measures (isolation-by-distance [IBD], Wright 1943).

Evolutionary relationships within each species were assessed with median-joining networks constructed with the program NETWORK version 4.5.1.0 (www.fluxus-engineering.com/network_terms.htm) using a median joining algorithm and default settings (as per Bandelt et al. 1999). Each haplotype was divided into sample site contributions as reflected by the pie diagrams.

Deviations from selective sequence neutrality were assessed by calculating Fu’s Fs (Fu 1997) for each population using ARLEQUIN; significance was tested with 99 999 permutations. Negative (and significant) Fs values indicate recent population expansion or directional selection (i.e., selective sweep). Each site (and species) was also fitted for the population parameter τ in order to estimate the time since the most recent population expansion (τ = 2μτ; as per Rogers and Hartvig 1992), where τ is the age of the population in generations and μ is the mutation rate per generation for the sequence (μ = number of bp · divergence rate within a lineage · generation time in years). A range of cyt b mutation rates are available from previous fish studies: 1% per million years (MY) within lineages (Bowen et al. 2001; Reece et al. 2010); 1.55% per MY within lineages or 1.55 × 10^8 mutations per site per year (Lessios 2008). Here we use both values to generate a range of divergence times. While generation time is unknown for our study species, we conditionally used the equation T = (α + ω)/2, where α is the
age at first reproduction and $\omega$ is the age at last reproduction (Pianka 1978). We therefore obtained a generation time of 3.25 years for both species based on existing life-history information (Roelofs and Silcock 2008). Although absolute time of expansion estimates should always be interpreted with caution, comparisons between these closely related taxa are likely robust to parameter approximations.

Past population dynamics of each butterflyfish species were also estimated with coalescent Bayesian skyline plots (BSP) using BEAST version 1.7.1 (Drummond et al. 2005). The BSP model generates a posterior distribution of effective population size through time using Markov Chain Monte Carlo (MCMC) sampling. Because the model assumes a single panmictic population, we analyzed the 2 species separately, using a randomized subset of sequences for C. ornatissimus ($N = 134$) to standardize sample size between species. The number of grouped intervals was set to 5 and the MCMC analysis was run for $3 \times 10^5$ generations (sampled every 1000 iterations), of which the first 10% was discarded as burn-in. The substitution model used was informed by prior runs in jModelTest for each species (see above). All analyses were repeated using different clock models (strict clock and relaxed clock) with uncorrelated rates drawn from a lognormal distribution. We therefore conducted 2 independent MCMC runs for each parameter combination to ensure convergence and congruent effective sample sizes. The median and corresponding credibility intervals of the BSP were depicted using Tracer version 1.5 (available at http://evolve.zoo.ox.ac.uk/software/).

The time of divergence between species can also be estimated in ARLEQUIN as the corrected pairwise sequence distance between species minus the corrected pairwise sequence distance within a species, with which we here refer to as the corrected sequence divergence ($d_{cor}$, Tamura and Nei 1993). This method will sometimes lead to an overestimate of divergence time, especially where effective population is large (Edwards and Beerli 2000; Crandall et al. 2008), but we are here interested in only whether the splitting of these lineages predates the onset of Plio-Pleistocene glaciations.

### Microsatellite Genotype Analysis

Each species was genotyped at 10 microsatellite loci, 8 of which (GenBank Accession numbers: HQ163780 to HQ163787) were developed using procedures described in DiBattista and Feldheim (2010). Two additional pairs of existing butterflyfish primers (GenBank Accession numbers: GQ891257 and GQ891265) were also used to genotype samples (Lawton et al. 2010) because both cross-amplified and were variable. PCR reaction mixes and cycling parameters followed DiBattista and Feldheim (2010), but annealing temperatures of 56°C were used for the latter.

PCR products labeled with different dye colors were pooled for genotyping and resolved using an ABI 3130XL Genetic Analyzer (Applied Biosystems) along with a fluorescently labeled internal size standard (LIZ-500; Applied Biosystem); allele sizes were assigned manually with GENEMAPPER version 3.7 (Applied Biosystems). For each locus, the mean number of alleles ($N_a$), observed ($H_o$) and expected ($H_e$) heterozygosities, departure from Hardy-Weinberg proportions (HWE), and linkage disequilibria (LD) were assessed with GENEPOP version 4.0 on the web (genepop.curtin.edu.au/; Raymond and Rousset 1995); $N_A$ was also standardized for sample size by rarefaction in FSTAT version 2.9.3 (A, allelic richness, Goudet 2001). The possible presence of null alleles, allelic dropouts, and excessive stutter peaks was tested with MICRO-CHECKER version 2.2.3 (Van Oosterhout et al. 2004) by adjusting significance levels for multiple comparisons (sequential Bonferroni correction). Genotyping error rate was also estimated by independently regenotyping randomly selected samples (i.e., 10% of all specimens per locus, for each species, see Hoffman and Amos 2005) and then comparing them to the initial genotype at each locus; PCR amplification or scoring errors were negligible in this case (< 1.35% for each individual locus). With few exceptions (C. meyeri, $N = 6$), specimens amplified for at least 5 of the 10 microsatellite loci (but most at all loci) and were therefore included in subsequent analyses.

Genotypic population structure was assessed separately for each species as described above: 1) AMOVA without group subdivisions, 2) pairwise population $F_{ST}$ estimation, and 3) Mantel tests for IBD correlating geographic distance with genetic differentiation. Moreover, to compensate for the downward bias in $F_{ST}$ associated with highly variable microsatellites, these values are corrected ($G_{ST}$ in Hedrick 2005) using SMOGD version 1.2.5 (Crawford 2010).

To define genetic clusters (i.e., populations) without a priori information on the geographical origin of specimens, we used a Bayesian approach implemented in STRUCTURE version 2.3.2 (Pritchard et al. 2000). STRUCTURE assigns individual multilocus genotypes to 1 or more clusters by minimizing deviations from Hardy-Weinberg and linkage equilibrium. STRUCTURE also allows information about the geographical origin of the samples to be included but does not bias the analysis with this information to indicate population differences. The most likely number of clusters in the dataset was identified based on the probability of $K = 1$ to $K = 6$ or $K = 1$ to $K = 8$ for C. meyeri and C. ornatissimus, respectively, averaged over 10 replicate runs with 1 000 000 MCMC repetitions each and a burn-in of 10 000 iterations. Correlated allele frequencies and admixed populations were assumed. The most likely value of $K$ was determined by plotting the mean Ln probability of the data (Ln P[D]) over multiple runs versus $K$ for each species. Identical results were obtained using the Delta K method of Evanno et al. (2005) to determine the most likely value of $K$ and so only the results of the former are presented here. Cluster plots were then generated using runs from the most likely $K$ in STRUCTURE.

Finally, to characterize relationships between mtDNA and microsatellite datasets within the present study and additional published studies on marine fish with pelagic larval stages, we ran separate Mantel correlation tests in zt version 1.1 (Bonnet and Van de Peer 2002) using pairwise $F_{ST}$ and $G_{ST}$ values reported for each species. This allowed us to assess
whether the mtDNA and microsatellite patterns were concordant within datasets.

Results

Molecular Characteristics

Cyt b sequence data from C. meyeri and C. ornatissimus revealed 34 and 41 haplotypes, respectively (range: 5 to 14 and 7 to 17 within-sites), with haplotype diversity ranging from $b = 0.29$ to 0.93 and $b = 0.59$ to 0.90 and nucleotide diversity ranging from $\pi = 0.0012$ to 0.0030 or $\pi = 0.0030$ to 0.0050 within species (Table 1). Nucleotide diversity was almost twice as large for C. ornatissimus versus C. meyeri, but may have been influenced by greater sampling effort in the former ($N = 296$ versus $N = 134$). The 2 most common haplotypes within each species were detected at almost every site.

Among the individuals scored for the 10 microsatellite loci (Appendix S1 in Supporting Information), the mean number of alleles per locus was $18.80 \pm 1.35$ SEM (range: 12 to 26 alleles) or $21.10 \pm 2.16$ SEM (range: 13 to 36 alleles), allelic richness was $5.77 \pm 0.18$ SEM (range: 4.87 to 6.41) or $11.35 \pm 0.85$ SEM (range: 7.46 to 15.46), and observed heterozygosity ranged from 0.67 (Cor21) to 0.95 (Cor28) and 0.73 (Cor39) to 0.91 (Cor28) for C. meyeri and C. ornatissimus, respectively. Few loci deviated from Hardy-Weinberg equilibrium (within-site comparisons for C. meyeri: 3 out of 60, $P < 0.011$; within-site comparisons for C. ornatissimus: 5 out of 80, $P < 0.010$) and no linkage disequilibrium was detected in 270 and 360 within-site comparisons (between loci) after correcting for multiple tests ($P > 0.035$ in all cases).

Evidence of null alleles was detected in only 6 out of 60 and 8 out of 80 within-site comparisons, although 2 of the loci were disproportionately represented. Cor21 and Cor16 tested significant for null alleles 60% or 38% of the time for C. meyeri and C. ornatissimus, respectively. We therefore ran all subsequent analyses excluding or including these loci to mitigate bias; our findings were no different between datasets, so we present results including all 10 microsatellite loci.

Population Structure Analysis

In all cases the microsatellite $F_{ST}$ values reported here are corrected for high heterozygosity as per Hedrick (2005). Analyses of molecular variance revealed low or no population structure for C. meyeri (mtDNA: $\Phi_{ST} = 0.073, P < 0.001$; microsatellites: $F_{ST} < 0.001$, N.S.), but higher structure for C. ornatissimus (mtDNA: $\Phi_{ST} = 0.101, P < 0.001$; microsatellites: $F_{ST} = 0.016, P < 0.001$). To ensure that the pattern of microsatellite structure (or lack therefore) was not being driven by a single locus, we repeated AMOVA analyses for each species 1 locus at a time, which gave consistent results across all loci (data not shown). Population pairwise tests ($\Phi_{ST}$ or $F_{ST}$) revealed that mtDNA haplotype frequencies were significantly different in only 3 out of 15 and 11 out of 28 comparisons for C. meyeri and C. ornatissimus, respectively (Table 2; Appendix S2 in Supporting Information). Microsatellite allele frequencies ($F_{ST}$), on the other hand, were never significantly different for C. meyeri, but significant in 18 out of 28 comparisons for C. ornatissimus (Table 2; Appendix S2 in Supporting Information). Pairwise differentiation was therefore consistently weak for C. meyeri, but much greater for C. ornatissimus.

Table 1  Sample size and molecular diversity indices for Chaetodon meyeri and Chaetodon ornatissimus based on mitochondrial DNA (cytochrome b) sequence data. Time since the most recent population expansion was calculated using a range of mutation rates (1–1.55% per million years within lineages, Bowen et al. 2001; Lessios 2008; Reece et al. 2010) and a generation time of 3.25 years for both species (see Methods)

<table>
<thead>
<tr>
<th>Collection locality</th>
<th>n</th>
<th>$H_N$</th>
<th>Time since expansion (Yrs)</th>
<th>Haplotype diversity (h ± SD)</th>
<th>Nucleotide diversity ($\pi$ ± SD)</th>
<th>Fu’s $F_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chaetodon meyeri</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Republic of Seychelles (SEY)</td>
<td>6</td>
<td>5</td>
<td>62900–97500</td>
<td>0.93 ± 0.12</td>
<td>0.0022 ± 0.0017</td>
<td>-2.61a</td>
</tr>
<tr>
<td>Diego Garcia (DIG)</td>
<td>36</td>
<td>13</td>
<td>71800–111200</td>
<td>0.60 ± 0.01</td>
<td>0.0016 ± 0.0012</td>
<td>-10.50</td>
</tr>
<tr>
<td>Cocos-Keeling Islands, Aus. (COC)</td>
<td>24</td>
<td>12</td>
<td>74300–115100</td>
<td>0.90 ± 0.05</td>
<td>0.0029 ± 0.0018</td>
<td>-6.43</td>
</tr>
<tr>
<td>Christmas Island, Aus. (XMA)</td>
<td>31</td>
<td>13</td>
<td>59900–92800</td>
<td>0.81 ± 0.07</td>
<td>0.0023 ± 0.0015</td>
<td>-8.23</td>
</tr>
<tr>
<td>Republic of Palau (PAU)</td>
<td>7</td>
<td>2</td>
<td>102600–159000</td>
<td>0.29 ± 0.20</td>
<td>0.0012 ± 0.0011</td>
<td>1.51</td>
</tr>
<tr>
<td>Republic of Kiribati (KIR)</td>
<td>30</td>
<td>14</td>
<td>72900–113100</td>
<td>0.90 ± 0.03</td>
<td>0.0030 ± 0.0019</td>
<td>-7.88</td>
</tr>
<tr>
<td>All samples</td>
<td>134</td>
<td>34</td>
<td>61500–95400</td>
<td>0.85 ± 0.02</td>
<td>0.0024 ± 0.0016</td>
<td>-27.79</td>
</tr>
<tr>
<td>Chaetodon ornatissimus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cocos-Keeling Islands, Aus. (COC)</td>
<td>32</td>
<td>11</td>
<td>177500–275100</td>
<td>0.83 ± 0.05</td>
<td>0.0045 ± 0.0027</td>
<td>-1.71</td>
</tr>
<tr>
<td>Christmas Island, Aus. (XMA)</td>
<td>32</td>
<td>12</td>
<td>206400–310600</td>
<td>0.90 ± 0.03</td>
<td>0.0050 ± 0.0029</td>
<td>-2.08</td>
</tr>
<tr>
<td>Republic of Palau (PAU)</td>
<td>37</td>
<td>17</td>
<td>197100–305600</td>
<td>0.86 ± 0.05</td>
<td>0.0041 ± 0.0025</td>
<td>-8.02</td>
</tr>
<tr>
<td>Moorea, French Polynesia (MOR)</td>
<td>33</td>
<td>14</td>
<td>59400–92000</td>
<td>0.89 ± 0.04</td>
<td>0.0036 ± 0.0022</td>
<td>-5.86</td>
</tr>
<tr>
<td>Nuku Hiva, Marquesas (NUK)</td>
<td>37</td>
<td>12</td>
<td>15400–23900</td>
<td>0.82 ± 0.05</td>
<td>0.0033 ± 0.0020</td>
<td>-3.64</td>
</tr>
<tr>
<td>Republic of Kiribati (KIR)</td>
<td>50</td>
<td>14</td>
<td>216500–335600</td>
<td>0.75 ± 0.06</td>
<td>0.0033 ± 0.0020</td>
<td>-4.57</td>
</tr>
<tr>
<td>Palmyra Atoll (PAL)</td>
<td>14</td>
<td>8</td>
<td>205200–318100</td>
<td>0.89 ± 0.06</td>
<td>0.0043 ± 0.0026</td>
<td>-1.83</td>
</tr>
<tr>
<td>Hawaiian Islands (HAW)</td>
<td>61</td>
<td>7</td>
<td>185600–287600</td>
<td>0.59 ± 0.05</td>
<td>0.0030 ± 0.0019</td>
<td>0.90</td>
</tr>
<tr>
<td>All samples</td>
<td>296</td>
<td>41</td>
<td>184700–286300</td>
<td>0.83 ± 0.02</td>
<td>0.0041 ± 0.0024</td>
<td>-24.53</td>
</tr>
</tbody>
</table>

*aNumbers in bold are significant, $P < 0.05$ (or $P < 0.02$ for Fu’s $F_S$ estimates, Fu 1997).

*bAbbreviations are as follows: $H_N$, number of haplotypes; Aus., Australia.
Table 2  Matrix of population pairwise \( \Phi_{ST} \) values (below diagonal) and standardized \( F_{ST} \) values (\( G'_{ST} \); Hedrick 2005; above diagonal), with associated \( P \)-values in parentheses, based on mitochondrial DNA (mtDNA) cytochrome \( b \) sequence data and microsatellite genotypes, respectively, from (a) Chaetodon meyeri (mtDNA, \( N = 134 \); microsatellites, \( N = 128 \)) or (b) Chaetodon ornatissimus (mtDNA, \( N = 296 \); microsatellites, \( N = 296 \)) sampled at sites across the Indo-Pacific region. All negative \( \Phi_{ST} \) values were adjusted to 0.

(a)

<table>
<thead>
<tr>
<th>Location</th>
<th>SEY</th>
<th>DIG</th>
<th>COC</th>
<th>XMA</th>
<th>PAU</th>
<th>KIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEY</td>
<td></td>
<td>0.71</td>
<td>6.05 ( \times 10^{-7} )</td>
<td>0.19</td>
<td>0.68</td>
<td>0.10</td>
</tr>
<tr>
<td>DIG</td>
<td>0.134* (0.02)</td>
<td></td>
<td>0.002 (0.35)</td>
<td>0.011 (0.55)</td>
<td>0.08 (0.68)</td>
<td>0.013 (0.06)</td>
</tr>
<tr>
<td>COC</td>
<td>0.016 (0.29)</td>
<td>0.026 (0.06)</td>
<td></td>
<td>0.72</td>
<td>0.47</td>
<td>0.039 (0.09)</td>
</tr>
<tr>
<td>XMA</td>
<td>0.133* (0.02)</td>
<td>0.161* (0.001)</td>
<td>0.040* (0.05)</td>
<td></td>
<td>0.92</td>
<td>4.50 ( \times 10^{-6} ) (0.25)</td>
</tr>
<tr>
<td>PAU</td>
<td>0.291 (0.05)</td>
<td>0.303* (0.001)</td>
<td>0.083 (0.06)</td>
<td>0.052</td>
<td></td>
<td>0.048</td>
</tr>
<tr>
<td>KIR</td>
<td>0.076 (0.09)</td>
<td>0.122* (0.001)</td>
<td>0.014 (0.19)</td>
<td>0.63</td>
<td>0.003 (0.36)</td>
<td></td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Location</th>
<th>COC</th>
<th>XMA</th>
<th>PAU</th>
<th>MOR</th>
<th>NUK</th>
<th>KIR</th>
<th>PAL</th>
<th>HAW</th>
</tr>
</thead>
<tbody>
<tr>
<td>COC</td>
<td></td>
<td>0.003</td>
<td>(&lt;0.001)*</td>
<td>0.043</td>
<td>(&lt;0.001)*</td>
<td>0.020</td>
<td>0.007</td>
<td>0.075</td>
</tr>
<tr>
<td>XMA</td>
<td>0</td>
<td>0.003</td>
<td>0.019</td>
<td>0.011</td>
<td>0.014</td>
<td>0.053</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAU</td>
<td>0.068</td>
<td>0.112</td>
<td>(&lt;0.001)*</td>
<td>(&lt;0.001)*</td>
<td>(&lt;0.001)*</td>
<td>(&lt;0.001)*</td>
<td>(&lt;0.001)*</td>
<td>(&lt;0.001)*</td>
</tr>
<tr>
<td>MOR</td>
<td>0.191</td>
<td>0.220</td>
<td>0.033</td>
<td>(&lt;0.001)*</td>
<td>(&lt;0.001)*</td>
<td>0.113 ( \times 10^{-4} )</td>
<td>0.003</td>
<td>0.101</td>
</tr>
<tr>
<td>NUK</td>
<td>0.141</td>
<td>0.181</td>
<td>0.007</td>
<td>0.002</td>
<td>0.33</td>
<td>0.03</td>
<td>0.22</td>
<td>0.01</td>
</tr>
<tr>
<td>KIR</td>
<td>0.144</td>
<td>0.183</td>
<td>0.004</td>
<td>0.011</td>
<td>0.008</td>
<td></td>
<td>4.42 ( \times 10^{-6} )</td>
<td>0.101</td>
</tr>
<tr>
<td>PAL</td>
<td>0.077</td>
<td>0.112</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td></td>
<td>0.164</td>
</tr>
<tr>
<td>HAW</td>
<td>0.021</td>
<td>0.043</td>
<td>0.113</td>
<td>0.271</td>
<td>0.211</td>
<td>0.192</td>
<td>0.151</td>
<td></td>
</tr>
</tbody>
</table>

* Site abbreviations are described in Table 1.

b Significant values are indicated in bold: * for \( P \leq 0.05 \) and † for \( P \leq 0.013 \) (corrected as per Narum 2006).

Historical Demography

Tests for \( cyt b \) neutrality revealed negative and significant Fu’s \( F_{S} \) values in 5 out of 6 sites considered for \( C. meyeri \) (Fu’s \( F_{S} = -10.50 \) to 1.51), and 4 out of 8 \( C. ornatissimus \) sampling sites (Fu’s \( F_{S} = -8.02 \) to 0.90; Table 1). The range of mutation rates and \( \tau \) values yielded estimates of time since last population expansion for each sampling site and species, with \( C. meyeri \) (61 500 to 95 400 years) expanding more recently than \( C. ornatissimus \) (184 700 to 286 300 years; Table 1). Bayesian skyline tests did not recover significant population size changes for either butterflyfish species based on the wide upper and lower 95% confidence intervals and relatively low effective sample size values (< 200 in all cases; Appendix S4 in Supporting Information). Despite this uncertainty, our findings based on demographic expansion analysis indicate historically recent population expansion in \( C. meyeri \) throughout its range, and much older expansion events for \( C. ornatissimus \). Given that our expansion estimates here reflect only the most recent genetic bottleneck in each region and not the timing of lineage divergence for these closely related taxa (see Craig et al. 2010), we considered the

Historical Demography

Tests for cyt b neutrality revealed negative and significant Fu’s \( F_{S} \) values in 5 out of 6 sites considered for \( C. meyeri \) (Fu’s \( F_{S} = -10.50 \) to 1.51), and 4 out of 8 \( C. ornatissimus \) sampling sites (Fu’s \( F_{S} = -8.02 \) to 0.90; Table 1). The range of mutation rates and \( \tau \) values yielded estimates of time since last population expansion for each sampling site and species, with \( C. meyeri \) (61 500 to 95 400 years) expanding more recently than \( C. ornatissimus \) (184 700 to 286 300 years; Table 1). Bayesian skyline tests did not recover significant population size changes for either butterflyfish species based on the wide upper and lower 95% confidence intervals and relatively low effective sample size values (< 200 in all cases; Appendix S4 in Supporting Information). Despite this uncertainty, our findings based on demographic expansion analysis indicate historically recent population expansion in \( C. meyeri \) throughout its range, and much older expansion events for \( C. ornatissimus \). Given that our expansion estimates here reflect only the most recent genetic bottleneck in each region and not the timing of lineage divergence for these closely related taxa (see Craig et al. 2010), we considered the
average corrected nucleotide difference between our study species. Sequence divergence between *C. ornatissimus* and *C. meyeri* haplotypes \( d_{corr} = 0.079 \) is consistent with divergence approximately 3 to 4 MY ago (based on a mutation rate of 2% per MY between lineages; Bowen et al. 2001; also see Hsu et al. 2007). Sequence divergence within each species, however, was low (*C. ornatissimus*, \( d = 0.002 \); *C. meyeri*, \( d = 0.004 \)). Comparisons with the only other butterflyfish included in their sister group (*Chaetodon reticulatus*, Cuvier 1831; GenBank Accession number, FJ167700) yielded similar levels of genetic differentiation (*C. ornatissimus*, \( d = 0.067 \); *C. meyeri*, \( d = 0.082 \)).

**Concordance between Marker Classes**

The correlation of pairwise mtDNA \( \Phi_{ST} \) and microsatellite \( F_{ST} \) estimates was not significant for *C. meyeri* \( (r = 0.017, P = 0.42) \), but was highly significant for *C. ornatissimus* \( (r = 0.44, P = 0.021) \). Taken together with the results of 12 other marine fish studies that provided pairwise values for both \( \Phi_{ST} \) and \( F_{ST} \), the correlation between the 2 datasets was not significant in 8 out of 14 cases (Table 3).

**Discussion**

We examined the influence of historical biogeographic barriers on the genetic structure of two closely related Indo-Pacific butterflyfish taxa. Few studies have previously accomplished this task with species that share almost identical habitat preferences, life-history characteristics, and have overlapping ranges (Lieske and Myers 1994; Rocha et al. 2002). Analyses of both mitochondrial and microsatellite DNA revealed divergent patterns of genetic structure and population history between *C. meyeri* and *C. ornatissimus*. Comparisons between intraspecific datasets were also informative for addressing the potential limitations of genetic markers.

Prior to dissecting these results, we address the classification of *C. meyeri* and *C. ornatissimus* as sister taxa. A recent phylogeny of butterflyfishes supports this relationship with bootstrap values of 84% to 98% (Bellwood et al. 2010). An independent assessment of Chaetodontid phylogenetics by Fessler and Westneat (2007), however, places a third closely related species, *C. reticulatus*, as sister to *C. meyeri*. Introggression might explain this discrepancy as *C. meyeri* and *C. reticulatus* are sympatric over most of their range and hybridize extensively (Senou et al. 2006; Hsu et al. 2007; R.L. Pyle and J. E. Randall pers. comm.). Indeed, the *C. meyeri*/*C. reticulatus* in Fessler and Westneat (2007) had almost identical mtDNA, invoking the possibility of introgression between those 2 species. Fessler and Westneat (2007) also used an alternative suite of genes relative to Bellwood et al. (2010), and so this may represent an instance where the fit between a genealogy and the actual relationship among taxa is sensitive to the genetic markers chosen (e.g., Edwards et al. 2007). Regardless of the phylogenetic resolution, these branch orders do not detract from the characterization of our 2 study species as closely related, overlapping in distribution,
Table 3  Intraspécific genetic structure in marine fishes with pelagic larvae based on mitochondrial DNA (mtDNA) and microsatellite data ($F_{ST}$ or $F_{CT}$). Species common (and scientific) name, mtDNA gene and the number of microsatellite loci used, the number of significant pairwise population comparisons, sample size ($N$, mtDNA/microsatellite), Mantel correlation between pairwise population estimates for both markers, and references are listed here.

<table>
<thead>
<tr>
<th>Species</th>
<th>mtDNA gene</th>
<th>Microsatellite loci</th>
<th>Pairwise comp. (mtDNA)</th>
<th>Pairwise comp. (Microsatellite)</th>
<th>N</th>
<th>Mantel test</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scrawled Butterflyfish (Chaetodon meyeri)</td>
<td>Cyt b</td>
<td>10</td>
<td>3 out of 15</td>
<td>0 out of 15</td>
<td>134/128</td>
<td>r = 0.017</td>
<td>This study</td>
</tr>
<tr>
<td>Ornate Butterflyfish (Chaetodon ornatusimus)</td>
<td>Cyt b</td>
<td>10</td>
<td>11 out of 28</td>
<td>18 out of 28</td>
<td>296/296</td>
<td>r = 0.44</td>
<td>This study</td>
</tr>
<tr>
<td>Oval Butterflyfish (Chaetodon lunulatus)</td>
<td>D-loop</td>
<td>12</td>
<td>4 out of 10</td>
<td>7 out of 10</td>
<td>135/231</td>
<td>r = 0.86</td>
<td>Lawton et al. 2011</td>
</tr>
<tr>
<td>Chevron Butterflyfish (Chaetodon trifascialis)</td>
<td>D-loop</td>
<td>11</td>
<td>5 out of 10</td>
<td>0 out of 10</td>
<td>135/209</td>
<td>r = 0.81</td>
<td>Lawton et al. 2011</td>
</tr>
<tr>
<td>Roman Seabream (Chrysipterus laticeps)</td>
<td>Cyt b</td>
<td>7</td>
<td>0 out of 10</td>
<td>1 out of 10</td>
<td>107/179</td>
<td>r = 0.51</td>
<td>Teske et al. 2010</td>
</tr>
<tr>
<td>Patagonian toothfish (Disostichus eleginoides)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rogers et al. 2006</td>
</tr>
<tr>
<td>Blue Threadfin (Eutetractus tetraactylus)</td>
<td>COI</td>
<td>5</td>
<td>13 out of 15</td>
<td>13 out of 15</td>
<td>240/288</td>
<td>r = 0.90</td>
<td>Horne et al. 2011</td>
</tr>
<tr>
<td>Humped Rockcod (Gauthieria gibberifrons)</td>
<td>D-loop</td>
<td>9</td>
<td>0 out of 15</td>
<td>0 out of 15</td>
<td>162/164</td>
<td>r = 0.33</td>
<td>Matschiner et al. 2009</td>
</tr>
<tr>
<td>Striped Sea Bream (Lithognathus mormyrus)</td>
<td>D-loop</td>
<td>9</td>
<td>5 out of 21</td>
<td>18 out of 21</td>
<td>143/401</td>
<td>r = 0.41</td>
<td>Sala-Bozano et al. 2009</td>
</tr>
<tr>
<td>Blueeye Soldierfish (Myripristis berndti)</td>
<td>Cyt b</td>
<td>8</td>
<td>9 out of 45</td>
<td>34 out of 45</td>
<td>343/345</td>
<td>p = 0.90</td>
<td>Muths et al. 2011</td>
</tr>
<tr>
<td>Crimson Jobfish (Pristipomoides filamentosa)</td>
<td>Cyt b</td>
<td>11</td>
<td>10 out of 45</td>
<td>10 out of 45</td>
<td>479/417</td>
<td>r = 0.63</td>
<td>Gaither et al. 2011b</td>
</tr>
<tr>
<td>Black-spot sea bred (Pagellus bogaraveo)</td>
<td>D-loop</td>
<td>10</td>
<td>5 out of 10</td>
<td>1 out of 10</td>
<td>350/350</td>
<td>r = 0.12</td>
<td>Stockley et al. 2005</td>
</tr>
<tr>
<td>Bluehead Wrasse (Thalassoma bifasciatum)</td>
<td>D-loop</td>
<td>5</td>
<td>0 out of 15</td>
<td>0 out of 15</td>
<td>264/267</td>
<td>r = 0.33</td>
<td>Haney et al. 2007</td>
</tr>
<tr>
<td>Yellow Tang (Zebrasoma flavescens)</td>
<td>Cyt b</td>
<td>14</td>
<td>49 out of 171</td>
<td>70 out of 171</td>
<td>654/654</td>
<td>r = 0.74</td>
<td>Eble et al. 2011a</td>
</tr>
</tbody>
</table>

ecologically identical to the limits of current resolution, and hybridizing at the junction of the Indian and Pacific Oceans.

Marine Biogeography

Marine biogeographic barriers influence historical connectivity in different ways. For example, a number of reef fish species show evidence of restricted gene flow within and between Indo-Pacific ocean basins (Acanthurus triostegus, Planes and Fauvelot 2002; genus Hippocampus, Lourie et al. 2005; Pomacentridae and Labridae, Drew et al. 2008; Dascyllus trimaculatus, Leray et al. 2010; Lutjanus fulvus, Gaither et al. 2010; Zebrasoma flavescens, Eble et al. 2011a), whereas patterns of genetic structure in other taxa tend to highlight the divergence of gene flow across thousands of kilometers (Myripristis berndti, Craig et al. 2007; Centropyge loriculus, Schultz et al. 2007; genus Naso, Horne et al. 2008; Lutjanus kasmira, Gaither et al. 2010; Acanthurus nigrofuscus, DiBattista et al. 2011; Acanthurus nigrofuscus, Eble et al. 2011b). Widespread dispersal of marine fauna following speciation events, and subsequent mixing of sister taxa, complicates interpretation of these patterns in the Indo-Pacific.

Butterflyfishes (family Chaetodontidae) may be particularly sensitive to the effects of historical marine barriers. 9 of 31 Chaetodontid species complexes have range partitions at the IPB (Blum 1989), and another study found concordant phylogenetic patterns among closely related butterflyfish species within the Indonesian archipelago (McMillan and Palumbi 1995). Moreover, obligate corallivores such as C. meyeri and C. ornatusimus, as well as other butterflyfish species, are highly dependent on coral reef ecosystems to meet habitat requirements (Öhman et al. 1998). Historical events that reduce both reef area and reef quality may more severely impact these ecologically specialized fish when compared to generalist species (Bellwood et al. 2010; Lawton et al. 2011).

We first evaluate population genetic divergence between Indian and Pacific Ocean sites within the 2 butterflyfish species, which was weak (or absent) for C. meyeri (Table 2). Higher differentiation between the western Indian Ocean (Seychelles and Diego Garcia) and the rest of the Indo-Pacific was expected given the genetic uniqueness of this subregion (Ridgway and Sampayo 2005; Eble et al. 2011b; Gaither et al. 2011a) and affinity to east African fauna (Spalding et al. 2007). In the present study, Diego Garcia in the central Indian Ocean was nearly equally differentiated from Seychelles to the west ($F_{ST} = 0.13$) and Christmas Island to the east ($F_{CT} = 0.16$) so that no clear affiliation emerged for this central Indian Ocean location. For C. ornatusimus, there were strong genetic breaks between the eastern Indian/West Pacific, Central Pacific, and

624
Hawaii (Table 2; Figure 2). The distribution of this species (although not the genetic partition) is consistent with a scenario of restricted dispersal at the IPB during the Pleistocene due to sea level changes, emergent land bridges, and altered oceanography (Haq et al. 1987).

The divergence of Hawaiian *C. ornatus* is not surprising given the genetic distinctiveness of Hawaiian fishes relative to other locations in the Pacific Ocean (Bernardi et al. 2001; Planes and Fauvelot 2002; Bay et al. 2004; DiBattista et al. 2011). Hawaii was strongly differentiated from even the nearest Central Pacific location, Palmyra in the northern Line Islands (mtDNA: $\Phi_{ST} = 0.15$; microsatellites: $F_{ST} = 0.075$, both $P < 0.001$). Such divergence in *C. ornatus* is coupled with the absence of *C. meyeri* in the Hawaiian biogeographic province. Based on the isolation of the Hawaiian island chain, colonization events are probably rare, and so peripheral isolation is the most likely mode of divergence from ancestral populations. Despite genetic differentiation of Hawaiian fish however, these individuals still shared the 2 most common mtDNA haplotypes with fish from all other sampling sites (Figure 3), including those in the eastern Indian Ocean (~12 000 kilometers between Cocos-Keeling Islands and Hawaii). This does not necessarily reflect gene flow on a recent evolutionary time scale given that the most frequently occurring haplotypes are generally regarded as the oldest, although microsatellite data were similarly structured.

The lack of congruence in genetic structure between the 2 butterflyfish species is surprising not only due to their striking similarity in life-history traits but also because they hybridize within areas of overlap in the eastern Indian Ocean (Hobbs et al. 2009). What can explain the strong genetic structure observed for one butterflyfish species but not the other? One obvious possibility is that the 2 species differ in the duration of their planktonic larval stage (i.e., pelagic larval duration, PLD); although we feel that this is unlikely for 2 reasons. First, even though no direct estimates of PLD for these 2 species are available, estimates from other butterflyfish indicate that they are high overall and quite similar (Dępczyński and Bellwood 2006). Second, PLD is often a poor predictor of genetic structure in the marine environment, with many studies refuting a strong relationship between these 2 parameters (for review see Weersing and Toonen 2009).

An alternative explanation for the observed difference, and one that we favor, is that these species have responded differently to shared environmental, ecological, and evolutionary regimes. Reconstructing such events from mtDNA cyt $b$ variation indicate that both species show population expansion during the Pleistocene (see Table 1 and “star-like” phylogeny in Figure 3), after an initial divergence 3–4 MY ago ($d_{cov} = 0.079$). This supports the idea that coral reef ecosystems in the Indo-Pacific were particularly susceptible to population contraction and extinction events throughout most of this period. *C. meyeri* appears to have expanded more recently (61 525 to 95 364 years) than *C. ornatus* (184 679 to 286 252 years), reinforcing the conclusion that their evolutionary history may be drastically different. Lawton et al. (2011) similarly found that 2 species of butterflyfishes sampled in the Pacific Ocean had different levels of population structure and demographic history, which indicates that these discordant patterns may be widespread among the Chaetodontidae, as well as other reef taxa (Table 3, see also Rocha et al. 2002; Bowen et al. 2006; Carpenter et al. 2011; Toonen et al. 2011).

We propose that *C. meyeri* only recently invaded the Pacific Ocean from the Indian Ocean, or Pacific cohorts were more severely impacted by population contraction (or local extinction) events than *C. ornatus*. Indeed, nucleotide diversity is twice as high for *C. ornatus* versus *C. meyeri*, which indicates a much longer and more stable population history in the former. Moreover, despite both species showing similar numbers of alleles at microsatellite loci, the allelic diversity was much higher for *C. ornatus* than *C. meyeri*. Extinction-recolonization processes therefore have the potential to mask phylogeographic signals associated with biogeographic barriers in the marine environment. The fact that *C. meyeri* has not colonized sites in the Central Pacific such as French Polynesia, Hawaii, and parts of the Line Islands, or perhaps did so at one time but has since been extirpated, supports such assertions. Thus, despite uncertainty between alternative methods for dating (also see Crandall et al. 2012), our study appears to indicate that population contraction (or extinction) events may be as important as colonization events in unraveling evolutionary history (Liu et al. 2006; Hsu et al. 2007).

Concordance between Marker Classes

The coapplication of mtDNA and nuclear microsatellites serves to illuminate inheritance dynamics responsible for the observed phylogeographic patterns. Just as sampling multiple codistributed species may afford replication for inferring common historical events, multiple genetic markers serve a similar purpose within species (Avise 2000; Eytan and Hellberg 2010). Microsatellite markers have been widely applied in recent years due to their ability to detect fine-scale genetic structure in taxa with large population sizes and high levels of gene flow. All population estimates based on microsatellite loci for *C. meyeri*, however, were essentially zero, but mtDNA showed some significant genetic distinctions (albeit weak) for Indian Ocean sites. The correlation of mtDNA $\Phi_{ST}$ and microsatellite $F_{ST}$ estimates (Table 3) was not significant for *C. meyeri* ($r = 0.017$, $P = 0.42$), but was highly significant for *C. ornatus* ($r = 0.44$, $P = 0.021$). The former outcome may be attributed to low overall structure in this species, and yet the lack of a correlation between mtDNA and microsatellite data is emerging as a common phenomenon with troubling implications.

In light of the inconsistent results we obtained from molecular markers in this study, we examine this recurring phenomenon in other marine fishes. Coastal marine fishes are especially useful for examining discordance in the population genetics of mitochondrial and nuclear loci because several confounding factors can be eliminated. First, dispersal occurs primarily as pelagic eggs and larvae, long before sexual dimorphisms in morphology or behavior arise, eliminating sex-biased dispersal as an explanation for this
discordance (Pardini et al. 2001; Bowen et al. 2005; but see Miller-Sims et al. 2008). Second, effective population sizes (\(N_e\)) are usually very large (typically \(10^4\) to \(10^6\)) in coastal marine fishes distributed across thousands of km. In these cases, unequal genetic drift due to the 4-fold difference in \(N_e\) between mtDNA and nuclear genes (Birky et al. 1983) will probably not be sufficient to overwhelm other forces acting on the genetic architecture of populations. Third, chromosomal sex determination predominates in this group (XY or ZW, Devlin and Nagahama 2002), so that biased sex ratios are unlikely to contribute to the observed patterns. Although hermaphroditism is common in reef fishes, it has never been recorded in butterflyfish, and would only act to further decrease the disparity in \(N_e\) between mitochondrial and nuclear loci.

Comparisons among published fish surveys that used both mtDNA and microsatellite markers reveal that only 6 out of the 14 species showed more genetic structure (as indicated by the number of significant pairwise comparisons) with microsatellite data, and of the other 8, 4 showed more population structure with mtDNA (Table 3). Another way in which we assessed the incongruence between mtDNA and microsatellite data was with Mantel partial correlation tests within each study. Although this is a weak statistical test in general, only 6 of 14 datasets had a significant correlation between pairwise mtDNA \(\Phi_{ST}\) and microsatellite \(F_{ST}\) estimates. This is surprising because based on the factors outlined above, and under assumptions of neutrality and mutation-drift equilibrium, both classes of markers should show similar patterns (but not magnitude, Larsson et al. 2009) of population differentiation.

What can explain the lack of concordance between molecular markers in general? One obvious explanation is that even species with large \(N_e\) may not be in mutation-drift equilibrium, in which case \(F_{ST}\) estimators are subject to bias. Indeed, following a disturbance such as an expansion event, mtDNA should return to equilibrium more quickly and show greater structure in general than nuclear loci (Crow and Aoki 1984). This idea is supported by the fact that tests for IBD signatures in the present study were nonsignificant for \(C.\) meyeri but approached significance for \(C.\) ornatissimus at mtDNA loci only, indicating that \(C.\) ornatissimus may be closer to equilibrium than \(C.\) meyeri (also see Liu et al. 2006).

Other possible explanations for the lack of concordance between molecular markers in our study include the potential for natural selection to drive differences between marker types, although selective sweeps for mtDNA are apparently rare (Mc cusker and Bentzen 2010; Karl et al. 2012). At least part of the explanation is linked to sample size, as the much higher diversity in microsatellites requires larger sample sizes to characterize allele frequencies. Another potential explanation is that the mtDNA analysis accounts for mutational distance between haplotypes while the microsatellite analysis uses allele frequencies only. This explanation is inadequate, however, because mtDNA haplotype frequencies (without mutational distance) provide nearly identical results (data not shown). Pairwise \(\Phi_{ST}\) estimates for mtDNA are also corrected for sample size, whereas pairwise \(F_{ST}\) estimates for microsatellite loci are not, although we have included studies with equal and unequal sample sizes for comparison (Table 3). Two final possibilities are that: 1) increased allelic homoplasy at microsatellite loci may mask genetic differentiation over long periods of time or when genetic differentiation is weak (O’Reilly et al. 2004), and 2) recent gene flow may erase genetic signatures at microsatellite loci more rapidly than mtDNA loci. Given that the reason for this discrepancy in the ability to detect genetic structure is not clear, and mtDNA and microsatellite results are not correlated in more than half of surveyed reef fishes (Table 3), we endorse the application of both nuclear and mitochondrial marker types where possible (Eytan and Hellberg 2010).

**Conclusions**

At the onset of this study we anticipated that \(C.\) meyeri and \(C.\) ornatissimus descended from an ancestral corallivore split into Indian and Pacific species by the IPB, with subsequent range expansion in both oceans. If this is the case, the 3 to 4 MY history of divergence has erased phylogeographic signatures within each species, possibly due to population and range expansions in the last 350 000 years. This finding indicates that while historical barriers may be important in driving diversification in these fishes, subsequent dispersal and extinction/recolonization events may obscure genetic signatures of these divisions. Given the similarities between these 2 closely related taxa in morphology, life history, and habitat preferences, in addition to their propensity to hybridize; it is likely that factors other than ecological differences (sensu Rocha et al. 2002) are responsible for contrasting phylogeographic patterns. Given that the only obvious difference between these species is centers of distribution in separate ocean basins, the crucial missing factor may be the biotic response of coral reefs to glacial cycles in different regions of the Indo-Pacific. Finally, our findings demonstrate higher population structure with mtDNA in \(C.\) meyeri and higher structure with microsatellites in \(C.\) ornatissimus. We discovered that this nonconcordance among genetic markers is widespread in reef fishes and probably other marine taxa, and cannot be easily explained by inheritance dynamics, sex-biased dispersal, or selection. Aside from the conventional cautions about relying on one or a few markers for population resolution, these findings provide additional reasons for casting the genomic net widely when resolving evolutionary history.

**Funding**

National Science Foundation (OCE-0453167, OCE-0929031 to B.W.B), NOAA National Marine Sanctuaries Program (MOA 2005-008/66882 to R.J.T.), HIMB-NWHI Coral Reef Research Partnership (NSMP MOA 2005-008/6682 to M.T.C), and Natural Sciences and Engineering Research Council of Canada (NSERC) postgraduate fellowship to J.D.D.
Acknowledgments

For specimens collection we thank Kim Andersen, Paul H. Barber, J. Howard Choo, G. Concepcion, Toby S. Daly-Engel, Joshua A. Drew, John L. Earle, Jeff Eble, Kevin Flanagan, Michelle R. Gaither, Brian D. Greene, Matthew Iacchei, Stephen A. Karl, Randall K. Kosaki, Carl G. Meyer, Yannis P. Papastamatiou, David Pence, Richard Pyle, Joshua Reece, D. Ross Robertson, Jennifer K. Schulz, Derek Skillings, Derek Smith, Zoltan Szabo, Kim Tenggardjaja, Bill Walsh, Ivar Williams, Jill P. Zamzow, and the crew of the R.V. Hi'ialakai. We thank Sue Taei at Conservation International, Graham Wragg of the RV Bounty Bay, the Government of Kiribati, including Takah Nakagawa, Delcy Tawani, and the Phoenix Island Protected Area who assisted with Kiribati collections. We also thank Robert Toonen, Seres Planes, Ben Victor, Jo-Ann Leong, Hawaii Department of Land and Natural Resources, Coral Reef Research Foundation and Papahanaumokuakea Marine National Monument, U.S. Fish and Wildlife Service, members of the ToBo lab for logistic support; and the Center for Genomics, Proteomics, and Bioinformatics at the University of Hawaii (Manoa Campus) for their assistance with genotyping. We thank Eric Cran and anonymous reviewers for their insightful comments on a previous version of this manuscript. This is contribution no. 1508 from the Hawai'i Institute of Marine Biology and no. 8713 from the School of Ocean and Earth Science and Technology.

References


Received March 12, 2012; Revised May 31, 2012; Accepted June 11, 2012

Corresponding Editor: C Scott Baker