MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 219–234 doi: 10.1111/j.1365-294X.2010.04942.x

The West Pacific diversity hotspot as a source or sink for new species? Population genetic insights from the Indo-Pacific parrotfish *Scarus rubroviolaceus*

J. M. FITZPATRICK,* D. B. CARLON,* C. LIPPE*‡ and D. R. ROBERTSON†

*Department of Zoology, University of Hawaii, Edmondson Hall 152, 2538 McCarthy Mall, Honolulu, HI 96822, USA,
†Smithsonian Tropical Research Institute, Box 0843-03092, Balboa, Republic of Panama

Abstract

We used a population genetic approach to quantify major population subdivisions and patterns of migration within a broadly distributed Indo-Pacific parrotfish. We genotyped 15 microsatellite loci in *Scarus rubroviolaceus* collected from 20 localities between Africa and the Americas. A STRUCTURE model indicates the presence of four major populations: Eastern Pacific, Hawaii, Central-West Pacific and a less well-differentiated Indian Ocean. We used the isolation and migration model to estimate splitting times, population sizes and migration patterns between sister population pairs. To eliminate loci under selection, we used BayeScan to select loci for three isolation and migration models: Eastern Pacific and Central-West Pacific, Hawaii and the Central-West Pacific, and Indian Ocean and the Central-West Pacific. To test the assumption of a stepwise mutation model (SMM), we used likelihood to test the SMM against a two-phase model that allowed mutational complexity. A posteriori, minor departures from SMM were estimated to affect ≤2% of the alleles in the data. The data were informative about the contemporary and ancestral population sizes, migration rates and the splitting time in the eastern Pacific/Central-West Pacific comparison. The model revealed a splitting time ~17 000 BP, a larger contemporary N_e in the Central-West Pacific than in the eastern Pacific and a strong bias of east to west migration. These characteristics support the Center of Accumulation model of peripatric diversification in low-diversity peripheral sites and perhaps migration from those sites to the western Pacific diversity hotspot.

Keywords: gene flow, IMa, microsatellites, peripatric speciation, stepwise mutation model

Received 8 September 2009; revision received 12 October 2010; accepted 15 October 2010

Introduction

The origin and maintenance of extraordinary species diversity found in the earth's tropics present substantial challenges to ecological and evolutionary theory. In the world's shallow oceans, species diversity attains a maximum within the Indo-Australian Archipelago (IAA), at the junction of the Indian and Pacific oceans, and declines to both the east and the west (Hughes *et al.* 2002). Three major models have been proposed to explain this classic pattern (reviewed by Barber *et al.* 2006): (i) 'Center of Origin', (ii) 'Center of Accumulation', and (iii) 'Center of Overlap'. The three models make contrasting predictions about the regional location of speciation, and the direction of species emigration. According to the Center of Origin model, speciation is concentrated in the IAA, and new species eventually diffuse eastward and westward towards Central Pacific and the Indian Ocean (Briggs 1999). Pleistocene isolation of the Pacific and Indian Oceans, driven by +100 m changes in sea level beginning ~2.6 Ma (Lambeck *et al.* 2002), provides a potential vicariance mechanism for recent speciation within the geographically complex IAA. In contrast, the Center of Accumulation model invokes peripatric speciation among peripheral

Correspondence: D. B. Carlon, Fax: 808 956 4722;
E-mail: carlon@hawaii.edu
‡Present address: BD Diagnostics, 2740 rue Einstein, Ste-Foy, Quebec G1P 4S4, Canada.

populations, followed by migration inwards towards the IAA (Jokiel & Martinelli 1992). In marine systems with high dispersal potential, peripatric diversification may be facilitated by the existence of numerous isolated island archipelagos around the margins of the Pacific. Finally, the Center of Overlap model predicts diversification within ocean basins, with faunas overlapping at the junction of basins because of changes in range limits (Woodland 1983). A variety of biogeographical and palaeontological data have been used previously to confront these models (Briggs 2003; Mora *et al.* 2003). Molecular data provide unique insights into the location and timing of historical diversification events and patterns of migration (Avise 2004), offering critical tests of the relative importance of these three hypotheses.

There is now clear phylogeographical support for the IAA as a centre of species origin (see reviews in: Williams *et al.* 2002; Barber *et al.* 2006; Crandall *et al.* 2008). Thus, it appears that geographical and/or environmental complexity of this region can contribute to population differentiation and speciation. On the other hand, there is also strong molecular evidence that support alternative models and implicate more complex processes in generating classical patterns of Indo-Pacific diversity. Many species show little or weak population structure across the IAA (Benzie 1999; Williams *et al.* 2002; Lessios *et al.* 2003; Crandall *et al.* 2008). Second, emerging species phylogenies of diverse and widespread tropical genera reveal that significant diversification can be associated with many geographical regions and that species age is not correlated with increasing range limits (Williams & Reid 2004; Meyer *et al.* 2005; Barber *et al.* 2006; Frey & Vermeij 2008; Malay & Paulay 2009). Lastly, there is evidence that pre-Miocene clades originated and diversified outside the Indo-West Pacific, then subsequently migrated towards the IAA region (Williams & Duda 2008; Frey & Vermeij 2008). Thus, processes that have generated the longitudinal diversity gradient across the tropics at one extreme may be unrelated to the recent past, or at another extreme they may reflect diversification and speciation events within the IAA that have been operating since the beginning of the Pleistocene.

Considering the Pleistocene time frame, there is a gap in understanding how species diversity accumulates within the Indo-West Pacific. From species phylogenies alone, there are limitations to inferring the geography of speciation from phylograms superimposed on maps of geographical ranges that may have changed considerably since speciation events (Losos & Glor 2003). Our approach to this problem is to focus on the intermediate phases of speciation by using broadly distributed tropical reef species and multilocus markers to understand the evolution of population structure. Specifically, we seek to understand the geography of potentially rapid population divergence and subsequent patterns of gene flow. We are aided by recent innovations in use of coalescent theory that permit detailed assessments of population splitting times, the dynamics of effective population sizes and biases in the direction/magnitude of migration among populations (Beaumont & Rannala 2004). Such population-level patterns offer tests of predictions of the Center of Origin vs. Center of Accumulation models, by indicating whether geographical variation in genetic diversity is driven by centripetal or peripatric processes and whether migration occurs away from or towards the IAA.

The parrotfishes (family Scaridae) are a pantropical group of charismatic reef herbivores that has diversified rapidly. The Scaridae is a monophyletic group with 90 described species in 10 genera (Parenti & Randall 2000). The most speciose genus *Scarus* (50 species) originated in the late Miocene (~10 Ma), then rapidly split into three major clades that diversified rapidly during the Pleistocene (Streelman *et al.* 2002; Alfaro *et al.* 2009). The biogeography of the Scaridae is similar to many tropical groups, with maximum diversity centred on the IAA, and decreasing diversity moving east and west (Fig. 1). Three Indo-Pacific parrotfish species are broadly distributed from Africa to the Pacific Americas: *Calotomus carolinus*, *Scarus ghoban* and *Scarus rubroviolaceus* (Robertson & Allen 2008). Currently there is only one published phylogeography of a widespread Indo-Pacific parrotfish, *Chlorurus sordidus* (Bay *et al.* 2004), but that species does not occur further east than the Central Pacific. In this study, we apply a population genetic approach to the conundrum of the West Pacific diversity hotspot by focusing on the broadly distributed *S. rubroviolaceus*. With data from 20 broadly distributed sampling localities and genotypes from 15 microsatellite loci (Carlon & Lippe 2007), we first estimate the number of populations and their geographical ranges using the model of STRUCTURE (Pritchard *et al.* 2000) and population graph theory (Dyer & Nason 2004). Second, we model historical splitting events between populations by fitting a series of isolation and migration models (IMa, Hey & Nielsen 2007) to sister populations. Finally, we use information on the geography of major populations, splitting events and patterns of gene flow to confront predictions of the Center of Origin vs. the Center of Accumulation model.

Materials and methods

Sampling and genotyping

Tissue samples were collected from individual fish between 1997 and 2008 from the Indian and Pacific

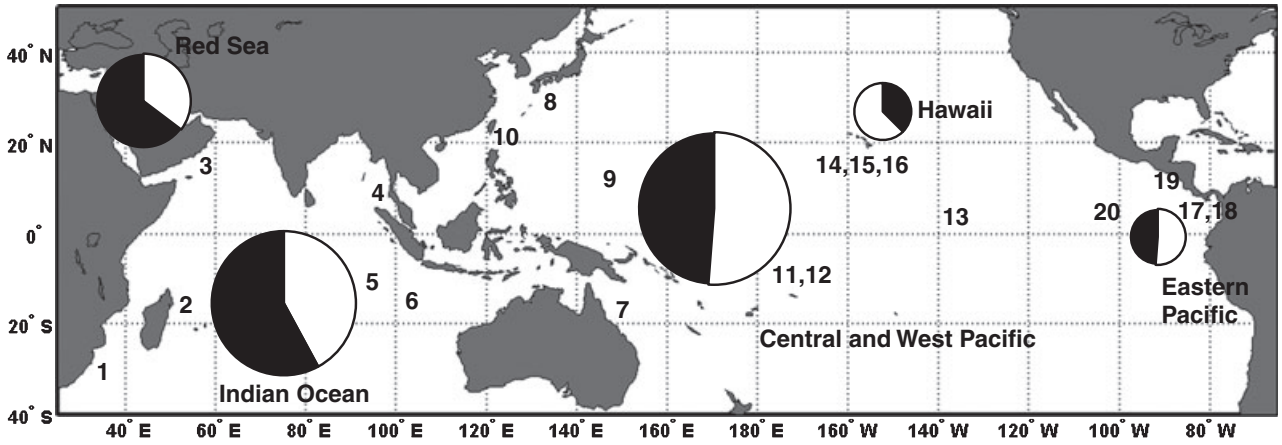

Fig. 1 Map of sampling localities including comparisons of parrotfish species diversity and endemism among five biogeographical regions. Size of pie charts is scaled to species diversity within regions, and black slices indicate the % of endemic species (data from Parenti & Randall 2000). Sampling localities: 1. Sodwana (n = 22); 2. Seychelles (n = 71); 3. Oman (n = 3); 4. Phuket (n = 24); 5. Christmas Island (n = 36); 6. Cocos (Keeling) Island (n = 8); 7. Great Barrier Reef (n = 13), 8. Okinawa (n = 2); 9. Rota Island, Northern Marianas Islands (n = 1); 10. Taiwan (n = 34); 11. Western Samoa (n = 46), 12. American Samoa (n = 18), 13. Ua Huka, Marquesas (n = 14); 14. Big Island, Hawaii (n = 42); 15. Maui, Hawaii (n = 24); 16. Oahu, Hawaii (n = 44); 17. Las Perlas Islands (n = 9); 18. Coastal Panama (n = 7); 19. Cocos Island (n = 51), 20. Clipperton Atoll (n = 6).

Oceans by spear, artisanal fish market sampling, larval sampling and nondestructive fin clips from sleeping fish. A total of 478 *Scarus rubroviolaceus* specimens were obtained from 20 localities throughout the Indian and Pacific Oceans (Fig. 1). Tissue was preserved in either a 20% DMSO salt solution or 95% ethanol before DNA extraction and amplification. Genomic DNA was extracted using the Qiagen DNeasy tissue extraction kit (Valencia, CA, USA) following the manufacturers instructions. Seventeen microsatellite loci developed for *S. rubroviolaceus* were amplified using PCR conditions described in Carlon and Lippé (Carlon & Lippe 2007). Two di-nucleotide loci (*Sru-A8* and *Sru-A9*) were excluded because of stutter problems and unreliable scoring. With the exception of one di-nucleotide locus (*Sru-A7*), all loci had tetra-nucleotide motifs and all show perfect repeat structure (*sensu* Estoup & Cornuet 1999). Samples were genotyped on an ABI 3700 sequencer and scored with GeneMapper software (Applied Biosystems, Inc.). Alleles were binned using ALLELOGRAM 2.2 (Carl Manaster, open source).

Summary statistics, Hardy–Weinberg expectations and linkage disequilibrium

Allele frequencies, diversity and observed/expected heterozygosity were calculated using MICROSATELLITE ANALYSER (MSA, v. 4.05 Dieringer & Schlotterer 2003). Within each sampling locality, we tested Hardy–Weinberg expectations (HWE) within loci and linkage disequilibrium (LD) between loci with GENEPOP 3.4 (http://genepop.curtin.edu.au/). To protect against excessive false positives in the large number of linkage tests, we used the Q-statistic, a measure of the false discovery rate (FDR) (Storey *et al.* 2003). Specifically, Q values measure the probability that a significant value is a false positive given alpha. We used the program QVALUE (http://www.genomics.princeton.edu/storeylab) to calculate Q values.

BayeScan

Methods of population structure and most coalescent-based models assume the molecular markers are neutral. To test this assumption, we used the program BayeScan (Foll & Gaggiotti 2008) to identify outlier loci. We used the three sister population pairs that we modelled with IMa and used the default settings for codominant data. We rejected neutrality at each locus if the $\log_{10}$ (Bayes Factor) >1.5. The results of BayeScans were used in two ways. First, we re-ran all population analyses (STRUCTURE, population graphs and F_{ST} comparisons) excluding loci that did not meet the criterion for neutrality. For IMa models, we excluded loci that failed neutrality (i.e. we could detect selection in BayeScans) in each specific population pair. It has been shown by Excoffier *et al.* (2009) that hierarchical population structure can increase the rate of false positives in the BayeScan approach. As there is evidence for structure between and within ocean basins in our data set, our Bayes Factors may be inflated by population phenomena unrelated to natural selection. This conservative approach to locus selection for IMa models will reduce the information content of the data, but more impor-

tantly it will minimize the potential effects of selection on model parameters.

Population structure

We used three methods to assess the population structure in the data. First, we used STRUCTURE (v 2.3, Pritchard *et al.* 2000) to estimate the number of populations (K) represented by all sampling localities and to determine the locations of major barriers to gene flow across the Indian and Pacific Oceans. We conducted 20 replicate runs for K values ranging from 1 to 7. All model runs used the admixture model with correlated allele frequencies, a burn-in of 10 000 steps and sampling runs of 100 000 steps. To determine which K was the best fit to the data, we used the statistic ΔK (Evanno *et al.* 2005). For graphical depiction of these results, we ran replicate runs through CLUMPP (Jakobsson & Rosenberg 2007) to match clusters and average Q values among runs, and DISTRUCT (Rosenberg 2004) to plot Q values from the resulting matrix. Second, we used population graphs (Dyer & Nason 2004) to visualize genetic correlations in the data, and test hypotheses about graph structure. We used all sampling localities (except Rota Island) as nodes and the program Genetic Studio (Dyer 2009) to estimate graphs and conduct statistical tests of connectivity among nodes. Lastly, we compared gene flow among populations determined by the STRUCTURE model of $K = 4$, and among all sampling localities using pairwise F_{ST}. For the among localities comparison, we excluded the single sample from Rota Island ($n = 1$). We used microsatellite analyzer (MSA) to calculate F_{ST} and the significance of F_{ST} by permuting 10 000 population pairs for each value. For the among locality comparison (19 localities total), we used Q values as in LD analyses to estimate the FDR and control for false positives.

Isolation and migration model

To model the history of isolation between populations, we fit isolation and migration models to sister population pairs with the program IMa (Hey & Nielsen 2007). The IMa model estimates the splitting time (t) of a single ancestral population with an effective population size (θ_A), into two derived populations that can have different sizes after splitting (θ_1 and θ_2) and asymmetrical migration (m_1 and m_2). Given the four populations indicated by STRUCTURE, we modelled three likely sister populations: (i) the Indian Ocean and the Central-West Pacific, (ii) the Central-West Pacific and the Hawaiian Islands, and (iii) the Central-West Pacific and the eastern Pacific. While geographical proximity dictates the sister pair of model (i), the sister pairs used in models (ii) and (iii) are based on the structure of oceanographic currents and the probability of larval migration among regions (Robertson *et al.* 2004). For example, the eastern Pacific is connected to the Central Pacific by two major current systems: the eastward moving Equatorial Countercurrent and the westward moving South Equatorial Current. Similarly, Hawaii is connected to the Central Pacific by predominant westward flow, making larval dispersal in the eastward direction much less likely. In model i (the Indian Ocean and the Central-West Pacific), we used sample groupings indicated by results from the STRUCTURE, $K = 4$ model. Specifically, samples from Cocos/Keeling and Christmas Island in the Indian Ocean (localities #5 and #6, Fig. 1) are included in the Central-West Pacific population.

With regard to the molecular markers, IMa makes three assumptions (i) that loci are not physically linked; (ii) that the loci are neutral with respect to natural selection; and (iii) that the loci fit a stepwise mutation model (SMM). Assumptions (i) and (ii) were tested with a linkage analysis and the outlier approach (e.g. BayeScan) described previously. In terms of mutational models, several pedigree studies in fishes (Jones *et al.* 1999; Steinberg *et al.* 2002) and a study in humans (Leopoldino & Pena 2003) have shown that mutations occur in a stepwise fashion at loci with perfect tetranucleotide structure, the same structure for 14 of 15 loci used here. Nonetheless, there are enough exceptions to the SMM within and between species (reviewed by Ellegren 2004) to warrant a closer look at mutational models in our markers. To test the SMM assumption, we used a likelihood approach implemented by MISAT (Nielsen 1997). The program MISAT calculates the likelihood of the data given a SMM, or alternatively a two-phase model (TPM) that allows a proportion of mutations (p) to involve changes greater than a single repeat. Model fit can then be statistically tested by the difference in likelihoods. Because this and other population tests of mutation models must simultaneously estimate θ (e.g. $4N_e u$), the outcomes are potentially confounded by departures from mutation–drift equilibrium within populations (Nielsen & Palsbøll 1999). To minimize the effects of such false positives, we compared the SMM to the TPM for each locus for samples collected from two different localities: the Island of Oahu, Hawaii and Western Samoa. By choosing two localities with reasonable sample sizes ($n \sim 50$ individuals) and different average θ values (Table 2, Oahu < -Western Samoa), we could test the SMM in populations with different demographic histories. The expectation is that while nonequilibrium conditions may increase (or decrease) the rate of false positives between populations, a particular mutation model will have the same

effect on allele size distributions regardless of the sample population. For each locus, Markov chains were run for 1 000 000 generations for SMM and TPM models, and p in the TPM was allowed to vary between 0.00001 and 0.5. We used likelihood ratio tests as described in Nielsen (1997) to determine which model better fit the data, and used Q values to correct for a total of 30 tests.

We conducted preliminary runs to identify reasonable priors for population sizes and migration rates. We chose the maximum values for the mutation-scaled splitting time (t) of 100 by assuming a mutation rate (μ) of 0.00094 (see below) and that maximum migration across the IAA was occurring ~150 000 BP at sea level heights comparable to present. Splitting time priors larger than this are not likely to be informative as all polymorphism is expected to coalesce within this time frame (t = 100 is >3× the largest geometric mean value of θ). To achieve suitable mixing, we used 200 Markov chains coupled by the Metropolis-Hastings algorithm. We used a geometric heating scheme to create slight differences in acceptance criteria among chains (heating values; $g1 = 0.99$; $g2 = 0.65$). These g values gave high swapping rates (>80%) for low numbered chains but enough swapping (>10%) in the highest numbered chains so that the parameter space was adequately explored. For each model we conduced two sampling runs, started from different random number seeds, with 2.5×10^6 burn-in steps, followed by $> 2 \times 10^6$ sampling steps with genealogies sampled every 100 steps. We used the L-mode of IMa to combine results from the two runs and estimate marginal and joint parameter distributions, and the nested model feature to examine specific hypotheses and reduced models when appropriate. This latter feature uses likelihood ratio tests to examine whether reduced models have greater likelihood. For example, a reduced model with fewer parameters would be a model where migration is symmetrical, or $m_1 = m_2$. Hypothesis testing proceeds by comparing twice the difference in model log likelihoods ($-2\hat{\Lambda}$) to χ^2 with k degrees of freedom, where k equals the difference in the number of parameters between the two models under consideration (Hey & Nielsen 2007). To convert the mutation-scaled IMa parameters to demographic quantities, we use a mutation rate (μ) from the literature and the strategy employed by Pavey *et al.* (2009) to estimate the geometric mean of μ among loci. We set μ of the tetra-nucleotide locus *Sru-B115* to the pipefish tetra-nucleotide locus *typh16* ($\mu = 0.00094$ per generation, Jones *et al.* 1999). This estimate should be reasonably close to *Sru-B115*, given that both loci are similar in maximum repeat length (*typh16* = 20, *Sru-B115* = 21), and that μ depends on repeat length in human tetra-nucleotide microsatellites (Leopoldino & Pena 2003). We used the μ scalar estimates from IMa to calculate the remaining mutation rates relative to *Sru-B115* and the geometric average of μ from all loci used in the specific model. With this geometric average μ, we calculated demographic equivalents of model parameters using equations given in the IMa documentation (http://genfaculty.rutgers.edu/hey/software). In Hawaii, Howard (2008) has shown the average age of maturity for *S. rubroviolaceus* is 5 years; we assumed this generation time to scale μ per generation to μ per year.

Results

Summary statistics, Hardy–Weinberg expectations and linkage disequilibrium

Between one and 25 alleles were detected at the 15 microsatellite loci within sampling localities (Appendix S1, Supporting information). Average allelic richness ranged from 2.8 to 12.3 with the lowest average allelic richness in the Eastern Pacific and Hawaii (range: 2.8–6.9). Average expected heterozygosity was also lowest in Hawaii and the Eastern Pacific (range: 0.575–0.628) and highest in the Central-West Pacific and Indian Ocean (range: 0.743–0.778). Locus *Sru-B115* and *Sru-D103* consistently deviated from HWE in almost all sampling localities (Appendix S1), and repeatable single-locus PCR failure at these two loci in some samples suggests the presence of null alleles, also confirmed with the program MICROCHECKER (v 2.2.3, van Oosterhout *et al.* 2004). As the estimated frequencies of the null alleles at these loci were low or moderate in the major populations indicated by the STRUCTURE, $K = 4$ model (Appendix S2, Supporting information), we included *Sru-B115* in all analyses and *Sru-D103* in some IMa models as described in the IMa section.

There was no evidence for physical linkage among any of the loci pairs. We detected significant LD at 34 loci pairs in four of 17 sampling localities with sample sizes >5 (Q values <0.01, Appendix S3, Supporting information). There were a total of 1432 possible tests. While three pairs of loci showed significant LD in two sampling localities (Cocos-Keeling Island and the Seychelles), they were not significant in any of the 15 other sampling localities. These patterns indicate that significant LD is driven by demographic phenomena within populations, rather than physical linkage.

BayeScan

BayeScan runs indicated strong evidence for selection in subsets of loci that depended on the population comparison (Table 1). One locus: *Sru-D4* showed strong

Table 1 BayeScan results from three population comparisons used for isolation and migration models. A Log_{10} (Bayes Factor) >1.5 is considered strong evidence for selection, indicated for each locus (within each comparison) by shading. Positive alpha values indicate directional selection, while negative values indicate stabilizing selection

	Central-West Pacific and Hawaii				Central-West Pacific and eastern Pacific				Central-West Pacific and Indian Ocean			
Locus	Prob	log_{10}(BF)	Alpha	F_{st}	Prob	log_{10}(BF)	Alpha	F_{st}	Prob	log_{10}(BF)	Alpha	F_{st}
A7	0.307	−0.354	0.028	0.219	0.389	−0.196	0.333	0.314	0.419	−0.142	0.443	0.108
B12	1.000	3.699	−1.323	0.072	0.577	0.135	−0.545	0.178	0.419	−0.141	−0.390	0.051
B115	0.264	−0.444	0.176	0.233	0.260	−0.455	0.123	0.271	0.231	−0.522	−0.082	0.065
B118	0.325	−0.317	0.174	0.242	0.368	−0.235	0.314	0.307	0.386	−0.201	0.331	0.097
C4	0.505	0.009	0.484	0.284	0.264	−0.445	−0.013	0.253	0.286	−0.396	−0.132	0.063
C11	0.627	0.225	0.599	0.312	0.378	−0.216	0.357	0.312	0.988	1.908	1.071	0.179
C105	0.343	−0.283	0.301	0.257	0.322	−0.323	0.257	0.296	0.973	1.553	1.096	0.183
C110	0.361	−0.249	0.353	0.264	0.361	−0.247	0.329	0.306	0.270	−0.432	−0.086	0.066
C121	0.270	−0.432	0.101	0.227	0.276	−0.419	0.031	0.260	0.316	−0.335	−0.207	0.060
C127	0.954	1.321	0.790	0.355	0.839	0.717	0.685	0.379	0.938	1.181	0.814	0.142
D4	1.000	1000.000	−1.581	0.055	1.000	1000.000	−1.706	0.065	1.000	1000.000	−1.221	0.022
D5	0.708	0.384	0.579	0.306	0.382	−0.208	0.371	0.311	0.586	0.151	0.508	0.103
D103	0.907	0.990	−0.695	0.124	0.978	1.644	−0.822	0.139	1.000	3.398	−1.144	0.023
D104	0.355	−0.259	−0.315	0.172	0.259	−0.457	0.002	0.255	0.524	0.041	0.490	0.103
D110	1.000	1000.000	−1.190	0.079	1.000	1000.000	−1.296	0.093	0.778	0.544	−0.557	0.042

evidence for stabilizing selection ($\alpha < -1.21$) in the three population comparisons, and two loci: *Sru-D103* and *Sru-D110* were outliers in two population comparisons, both with negative alpha values. In the total of six outlier loci detected in one or more population comparisons, *Sru-C105* and *Sru-C11* showed evidence for directional selection by positive alphas across the IAA in the Central-West Pacific vs. Indian Ocean comparison. The set of nine neutral loci used to repeat population structure estimates included the following: *Sru-A7*, *Sru-B115*, *Sru-B118*, *Sru-C4*, *Sru-C110*, *Sru-C121*, *Sru-C127*, *Sru-D5* and *Sru-D104*.

Stepwise mutation model

Of the 30 tests conducted, we could reject the SMM in favour of the TPM in 12 of them (Table 2). However, there were only three loci in which the SMM was rejected in favour of the TPM in both populations (*Sru-D4*, *Sru-D103* and *Sru-D110*). While we used the BayeScan results to select neutral loci for IMa models, tests of mutation models were suggested by reviewers and carried out after IMa runs were completed. There are two ways to determine the effects of departures from a SMM on our IMa results. Preferably, we could repeat IMa runs excluding all loci that did meet criteria for neutrality and fit a SMM. Because a run of each IMa model consumed >70 days of computer time, 350 days of computer time were invested in the results presented in this study, and repeating IMa runs is not a trivial undertaking. As an alternative, we estimated the effects of departures from a SMM a posteriori, based on the number of alleles evolving by non-SMM processes in each model. In terms of loci, only two loci that fit the TPM (TPM loci) passed through the BayeScan filter into IMa data sets, and no more than one TPM locus was included in any one IMa model (Table 3). Fortuitously, no TPM loci passed the BayeScan filter and were included in the Central-West Pacific and eastern Pacific model. In the remaining two models, we could estimate the effect of the TPM locus on the data by estimating the total proportion of the alleles that result from nonstepwise mutations. Recall that the TPM includes both stepwise mutations and a proportion of alleles evolving by mutations greater than one repeat (p). We can therefore estimate the fraction of alleles in the data set generated by nonstepwise mutations as: $\frac{p \times A}{N}$, where A equals the number of alleles sampled at the TPM locus, and N equals the total number of alleles sampled across all loci. For the Central-West Pacific and Hawaii model, the single TPM locus *Sru-D103* was included in the data. Given $P = 0.11$ (averaged over two populations, Table 2), $A = 24$ alleles and $N = 165$, we estimate 1.6% of the alleles in the data set result from nonstepwise mutations. Similarly, in the Central-West Pacific and Indian Ocean model, the TPM locus *Sru-D104* was included in the data set. The estimate of nonstepwise mutations in the data is 2.0%. Thus, we can conclude that a very small number of alleles generated by nonstepwise mutations were included in IMa model runs.

Table 2 Results of likelihood ratio tests comparing the fit of the stepwise mutation model (SMM) and two-phase model (TPM) to the data from two population samples. Shaded cells indicate rejection of the SMM in both sample populations

		SMM		TPM						
Locus	Population	θ	Log $L(\theta)$	θ	p†	Log $L(\theta)$	2 × Log L difference‡	P	Q§	Reject SMM?¶
A7	Oahu, Hawaii	4.05	−29.37	2.34	0.45	−26.86	5.02	0.025	0.058	ns
	Western Samoa	2.79	−24.64	1.17	0.45	−22.78	3.72	0.054	0.101	ns
B115	Oahu, Hawaii	4.45	−31.82	1.76	0.50	−25.59	12.47	0.000	0.002	**
	Western Samoa	24.71	−35.29	25.66	0.00	−35.26	0.06	0.801	0.935	ns
B118	Oahu, Hawaii	2.20	−19.55	1.92	0.00	−19.54	0.02	0.882	0.935	ns
	Western Samoa	2.31	−22.11	1.54	0.11	−21.43	1.37	0.242	0.403	ns
B12	Oahu, Hawaii	15.06	−39.22	15.48	0.00	−39.09	0.26	0.609	0.935	ns
	Western Samoa	32.52	−48.54	12.45	0.14	−42.68	11.72	0.001	0.002	**
C105	Oahu, Hawaii	1.77	−15.67	1.89	0.00	−15.68	0.01	0.903	0.935	ns
	Western Samoa	8.80	−26.62	5.91	0.10	−22.39	8.45	0.004	0.010	**
C11	Oahu, Hawaii	0.97	−10.84	0.92	0.00	−10.83	0.01	0.904	0.935	ns
	Western Samoa	4.56	−21.50	3.91	0.00	−21.58	0.15	0.699	0.935	ns
C110	Oahu, Hawaii	1.39	−10.22	1.14	0.00	−10.27	0.10	0.756	0.935	ns
	Western Samoa	14.65	−26.84	11.72	0.00	−27.57	1.46	0.226	0.399	ns
C121	Oahu, Hawaii	1.28	−20.77	0.84	0.48	−17.46	6.64	0.010	0.025	*
	Western Samoa	11.02	−30.30	11.44	0.00	−30.35	0.11	0.746	0.935	ns
C127	Oahu, Hawaii	3.14	−19.08	1.86	0.13	−16.98	4.19	0.041	0.084	ns
	Western Samoa	1.15	−2.04	1.09	0.00	−2.05	0.02	0.900	0.935	ns
C4	Oahu, Hawaii	2.16	−22.72	0.70	0.48	−17.52	10.40	0.001	0.004	**
	Western Samoa	10.54	−30.39	9.18	0.00	−30.30	0.19	0.659	0.935	ns
D103	Oahu, Hawaii	30.02	−45.50	6.52	0.12	−33.49	24.01	0.000	0.000	**
	Western Samoa	76.36	−74.02	55.30	0.10	−64.82	18.39	0.000	0.000	**
D104	Oahu, Hawaii	5.75	−24.98	6.02	0.00	−24.97	0.02	0.894	0.935	ns
	Western Samoa	14.77	−36.23	5.75	0.10	−31.13	10.20	0.001	0.004	**
D110	Oahu, Hawaii	56.91	−59.21	37.38	0.08	−51.25	15.93	0.000	0.000	**
	Western Samoa	83.12	−65.20	43.90	0.18	−50.94	28.52	0.000	0.000	**
D4	Oahu, Hawaii	95.08	−103.77	53.60	0.10	−82.47	42.60	0.000	0.000	**
	Western Samoa	83.53	−81.71	68.80	0.06	−72.32	18.78	0.000	0.000	**
D5	Oahu, Hawaii	1.31	−11.36	1.27	0.00	−11.35	0.01	0.935	0.935	ns
	Western Samoa	4.25	−26.33	2.10	0.48	−24.26	4.13	0.042	0.084	ns

†Estimated proportion of multistep mutations.
‡Twice the difference between log likelihoods is distributed as χ^2 with 1 d.f.
§The false discovery rate.
¶A test is judged significant and the SMM is rejected in favour of the TPM if $Q < 0.05$, * < 0.05, ** < 0.01.

Population structure

STRUCTURE analyses indicated the existence of three major populations: the Central-West Pacific, Hawaii and the eastern Pacific, with evidence for a weakly differentiated fourth population in the Indian Ocean (Figs 2 and 3a). When all loci were included, the slope of LnP(D) vs. K remains relatively steep after $K = 3$, resulting in a secondary 'peak' in delta K at $K = 4$ (Fig. 2a). This fourth population reflects structure between the Indian Ocean and Central-West Pacific population samples (Fig. 3a, lower two panels). In comparison, for the neutral data set, delta K peaks at $K = 3$ because the slope of LnP(D) vs. K levels off sharply after $K = 3$ (Fig. 2b). Visual inspection of Q values for $K = 4$ clearly shows that many individuals sampled at Cocos/Keeling and Christmas Islands in the Indian Ocean (localities 5 and 6) have high assignments to the Central-West Pacific population (light green) rather than the Indian Ocean (yellow). At Cocos/Keeling, where we sampled 36 individuals, 22 had Q values >0.80 for the Central-West Pacific population, while only six samples had Q values >0.80 for the Indian Ocean population. This same pattern was found at the Christmas Islands ($n = 8$), where four and two samples were assigned to the Central-West Pacific and Indian Ocean populations, respectively, using the same Q value criteria. These data indicate that this pair of localities occurs in a part of the East Indian Ocean where the Central-West Pacific and Indian Ocean populations mix.

Table 3 Loci used in three different isolation and migration (IMa) models that met neutrality assumption as determined by BayeScan. The mutation model for each locus (SMM, stepwise mutation model; TPM, two-phase mutation model) was determined by likelihood tests in two populations (see Table 2). Bold Xs indicate a loci that did not fit a SMM. CWP, Central-West Pacific, EP, eastern Pacific, IO, Indian Ocean

	IMa model			
Locus	CWP–Hawaii	CWP–EP	CWP–IO	Mutation model
A7	X	X	X	SMM
B12		X	X	SMM
B115	X	X	X	SMM
B118	X	X	X	SMM
C4	X	X	X	SMM
C11	X	X		SMM
C105	X	X		SMM
C110	X	X	X	SMM
C121	X	X	X	SMM
C127	X	X	X	SMM
D4				TPM
D5	X	X	X	SMM
D103	**X**			TPM
D104	X	X	X	SMM
D110			**X**	TPM
Total # loci	12	12	11	

The population graph of the data (Fig. 3b) is largely consistent with the STRUCTURE analyses, but reveals additional patterns in the data. The Hawaii localities define an independent graph of three nodes that is not connected to a larger graph of 15 nodes that includes geographical localities in the eastern Pacific, Central Pacific, West Pacific and Indian Ocean. Within this larger graph, there is significant structure between a subgraph defined by the eastern Pacific localities (G1, four nodes, five edges) and the subgraph defined by the Central Pacific, West Pacific and Indian Ocean localities (G2, 11 nodes, 20 edges). The probability that two connecting edges between subgraphs G1 and G2 occurs by chance is quite small (8.7×10^{-5}). The number of edges connecting Pacific Ocean nodes to Indian Ocean nodes is 11, and the probability that this many connections occur by chance is 0.18. Counter to the graph predicted by isolation by distance, the eastern Pacific subgraph (G1) is connected to the larger graph by the Oman node in the Indian Ocean, rather than nearby Central Pacific nodes. This could be explained by relatively low sample sizes for Oman and the two Panamanian localities (<10 individuals), however, removing the Oman samples from data and re-running the analysis resulted in a similar topology. Repeating the population graph with the neutral loci had minor effects on the topology within the Central Pacific and Indian Ocean and no effect on the independent Hawaii graph or the

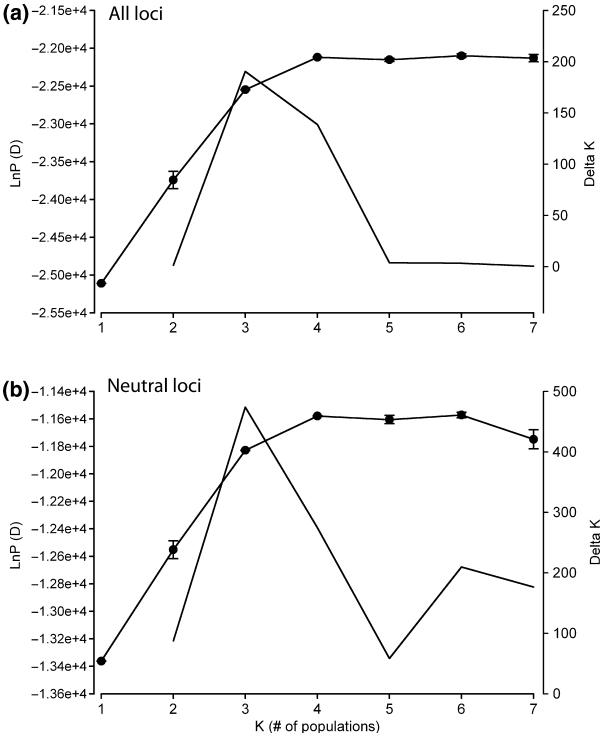

Fig. 2 Plots of LnP(D) (solid lines with symbols) and Delta *K* (solid lines) for STRUCTURE models using two data sets (a) 15 loci and (b) nine neutral loci. Error bars represent ±1 standard deviation. Nonvisible errors bars are smaller than the symbols.

number of edges connecting the eastern Pacific subgraph (G1) to the larger subgraph (G2).

Comparison of F_{ST} among samples grouped by the STRUCTURE, $K = 4$ model showed that genetic differentiation is highest among comparisons that include Hawaii and the eastern Pacific (all F_{ST} values >0.1, Table 4). In comparison, differentiation between the Indian Ocean and the Central-West Pacific was considerably less, with an F_{ST} of 0.029 from all loci and 0.036 from the nine neutral loci. If we grouped the Cocos-Keeling and Christmas Islands samples by strict geographical criteria (in the Indian Ocean) for this same comparison, rather than the STRUCTURE model, F_{ST} for all loci was slightly less = 0.021. All pairwise comparisons in this table were significant ($P < 0.001$). At smaller spatial scales, comparison of F_{ST} values calculated from neutral loci revealed significant differentiation within ocean basins (Table 5). In the Central-West Pacific population, Ua Huka (Marquesas) was significantly different from 7/8 possible within-population comparisons; and in the eastern Pacific population, Clipperton Atoll, located ~900 km from the mainland, was significantly different from the other three eastern Pacific sites: Cocos Island (350 km from the mainland) and Coastal Panama and the Pearl islands, which are on the continental shelf.

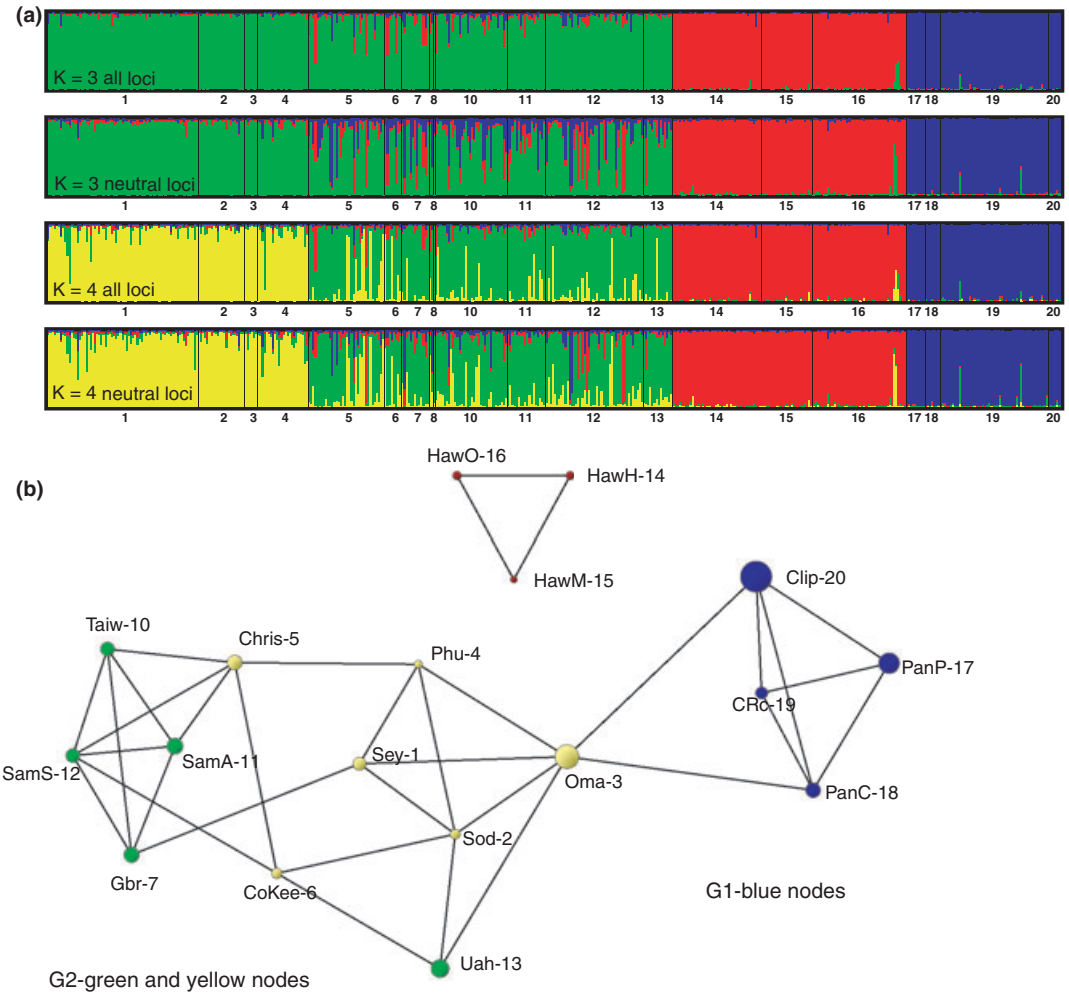

Fig. 3 (a) Structure plots for two *K* values and two data sets: all 15 loci vs. nine neutral loci. (b) Population graph depicting structure among localities using all loci. Node sizes are proportional to genetic diversity within samples, and edge length is scaled to genetic covariance among locations. Colour codes for (a) and (b): red = Hawaiian Islands, blue = Eastern Pacific, green = Central/West Pacific, yellow = Indian Ocean. For locality codes in (a) and (b) see Fig. 1. Rota Island (site 9, 1 sample) not included in this figure. The subgraphs G1 and G2 are compared statistically in the text.

IMa models

We could reliably estimate five of six parameters in the Central Pacific and eastern Pacific model. However, there was not enough information in the data to estimate several key parameters in the remaining two models (Central-West Pacific and Hawaii; Central-West Pacific and Indian Ocean) (Fig. 4). For those parameters with the strongest signals in the data (θ_1, θ_2, m_1, m_2), the two replicate runs of the Central-West Pacific and eastern Pacific gave similar maximum-likelihood estimates and distributions (Appendix S4, Supporting information). This suggests that the program was sampling from the stationary distribution and that parameter values are representative given the priors. We did not have enough time to replicate the model of the Central-West Pacific and Indian Ocean comparison for this study. Parameter estimates for this model should be treated with caution.

The Central-West Pacific and eastern Pacific model (Fig. 4a, Table 6) reveals that the contemporary effective population size of the eastern Pacific (N_2) is about 1/3 that of the Central Pacific (N_1) and that gene flow between these regions is highly asymmetric: on a demographic scale, the number of gene copies moving from East to West was ~12 copies per generation, while essentially no gene copies were moving into the eastern Pacific (Table 6, $2N_1m_1$ and $2N_2m_2$ respectively). Tests of two reduced models (Table 7) support this asymmetric scenario: a model with zero migration into

Table 4 Pairwise F_{ST} comparisons between four populations grouped by the STRUCTURE, $k = 4$ model. In this STRUCTURE model, the majority of the samples from two Indian Ocean localities: Cocos-Keeling and the Christmas Islands were assigned to the Central-West Pacific population (C-W Pacific) and were grouped in this population for these comparisons. Values above the diagonal are calculated from all 15 loci, while values below the diagonal are calculated from nine neutral loci. All pairwise comparisons using both all 15 and nine neutral loci were significant at $P > 0.001$

	Indian Ocean	C-W Pacific	Hawaii	Eastern Pacific
Indian Ocean		0.029	0.121	0.128
C-W Pacific	0.036		0.105	0.102
Hawaii	0.157	0.136		0.233
Eastern Pacific	0.140	0.094	0.273	

the eastern Pacific ($m_2 = 0$) does not significantly differ from the full model ($-2\hat{\Lambda} = 0.244$, d.f. = 1, $P = 0.621$), while a model with equal migration in both directions ($m_1 = m_2$) does differ from the full model ($-2\hat{\Lambda} = 45.122$, d.f. = 1, $P < 0.001$). Similarly, the effective population sizes of the Central and East Pacific are clearly different, as the model that sets $\theta_1 = \theta_2$ is rejected ($-2\hat{\Lambda} = 449.038$, d.f. = 1, $P < 0.001$). Not surprisingly, several models that involved differences between the ancestral population size (θ_A) and contemporary population sizes could not be rejected because of the broad distribution of posterior probabilities of θ_A.

Using the scaling approach to estimating mutation (Appendix S5, Supporting information), we find the geometric average $\mu = 0.00139$ locus/generation. Converting the mutation-scaled time estimate (t) to years gives a splitting time between the Central-West and East Pacific of 17 483 years before present.

While we did not convert mutation-scaled IMa parameters in the other models to demographic or time-calibrated equivalents, Hawaii also has a very recent splitting time ($t \sim 1.0$, Appendix S4, Supporting information). Assuming μ is similar to the Central-West Pacific and eastern Pacific model, the demographic colonization of Hawaii appears to be more recent than is the case with the eastern Pacific. The preliminary estimates of the Central-West Pacific and Indian Ocean model (Fig. 4c) suggest that the effective population size of the Indian Ocean is larger than the Pacific Ocean. However, the data do not allow estimation of splitting time and migration patterns between those areas.

Discussion

This microsatellite-based study has examined population structure and migration dynamics of a tropical reef fish with a range that spans the entire Indian and Pacific Oceans, including the remote Hawaiian Islands.

Table 5 Pairwise F_{ST} comparisons between 20 localities. Values above the diagonal are calculated from all 15 loci, while values below the diagonal are calculated from the set of nine neutral loci. Bold text indicates significant comparisons at $\alpha = 0.05$. The Q value for this alpha for both comparisons is <0.01. Numbers in the first row correspond to locations in Fig. 1

	1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20
	SEY	SOD	OMA	PHU	CHRS	CO/KE	AUS	JAP	TAI	SRA	SRS	UAH	HAWB	HAWM	HAWO	LASP	PAN	COC	CLIP
SEY		0.00	0.00	**0.01**	**0.02**	**0.04**	**0.04**	0.03	**0.03**	**0.02**	**0.04**	**0.05**	**0.12**	**0.12**	**0.11**	**0.14**	**0.13**	**0.13**	**0.14**
SOD	0.00		0.00	**0.00**	**0.03**	**0.05**	**0.04**	**0.04**	**0.03**	**0.03**	**0.04**	**0.05**	**0.14**	**0.14**	**0.13**	**0.14**	**0.14**	**0.13**	**0.15**
OMA	0.02	0.01		0.00	**0.02**	**0.05**	**0.06**	0.04	**0.04**	**0.03**	**0.03**	**0.03**	**0.17**	**0.16**	**0.16**	**0.17**	**0.14**	**0.14**	**0.18**
PHU	0.00	**0.00**	0.01		**0.02**	**0.03**	**0.03**	0.03	**0.04**	**0.02**	**0.05**	**0.05**	**0.13**	**0.13**	**0.12**	**0.16**	**0.15**	**0.15**	**0.16**
CHRS	**0.03**	**0.04**	**0.04**	**0.03**		0.01	**0.01**	0.01	0.00	0.00	**0.01**	**0.03**	**0.10**	**0.11**	**0.09**	**0.11**	**0.11**	**0.11**	**0.13**
CO/KE	**0.05**	**0.06**	**0.08**	**0.04**	0.01		**0.04**	0.02	**0.02**	0.02	**0.02**	**0.06**	**0.13**	**0.12**	**0.11**	**0.14**	**0.13**	**0.12**	**0.18**
AUS	**0.05**	**0.05**	**0.08**	**0.04**	**0.02**	**0.04**		0.02	**0.01**	0.00	**0.01**	**0.04**	**0.13**	**0.14**	**0.12**	**0.13**	**0.15**	**0.13**	**0.15**
JAP	0.04	**0.05**	0.06	0.02	0.01	−0.02	0.05		0.01	0.01	−0.03	0.06	**0.15**	**0.17**	**0.14**	**0.15**	**0.18**	**0.11**	**0.23**
TAI	**0.04**	**0.05**	**0.06**	**0.04**	0.00	**0.02**	**0.02**	0.02		0.00	**0.01**	**0.02**	**0.11**	**0.11**	**0.09**	**0.10**	**0.10**	**0.09**	**0.12**
SRA	**0.03**	**0.04**	**0.06**	**0.03**	−0.01	0.01	0.00	0.03	0.00		0.00	**0.02**	**0.11**	**0.11**	**0.10**	**0.12**	**0.12**	**0.11**	**0.13**
SRS	**0.04**	**0.06**	**0.05**	**0.06**	**0.01**	0.03	0.02	−0.05	**0.01**	0.00		**0.02**	**0.12**	**0.13**	**0.11**	**0.12**	**0.11**	**0.11**	**0.13**
UAH	**0.06**	**0.05**	**0.04**	**0.05**	**0.04**	**0.06**	**0.04**	**0.10**	**0.03**	**0.03**	**0.03**		**0.15**	**0.17**	**0.15**	**0.15**	**0.17**	**0.14**	**0.17**
HAWB	**0.15**	**0.19**	**0.25**	**0.17**	**0.14**	**0.17**	**0.18**	**0.21**	**0.15**	**0.16**	**0.17**	**0.21**		0.00	0.00	**0.25**	**0.26**	**0.23**	**0.26**
HAWM	**0.15**	**0.19**	**0.24**	**0.17**	**0.14**	**0.17**	**0.18**	**0.23**	**0.15**	**0.15**	**0.17**	**0.21**	0.00		0.00	**0.27**	**0.26**	**0.23**	**0.27**
HAWO	**0.13**	**0.16**	**0.22**	**0.15**	**0.11**	**0.14**	**0.16**	**0.18**	**0.12**	**0.13**	**0.14**	**0.19**	0.00	−0.01		**0.25**	**0.25**	**0.22**	**0.25**
LASP	**0.16**	**0.17**	**0.22**	**0.17**	**0.10**	**0.12**	**0.13**	**0.15**	**0.09**	**0.13**	**0.12**	**0.16**	**0.31**	**0.32**	**0.29**		0.00	0.00	0.02
PAN	**0.12**	**0.14**	**0.13**	**0.13**	**0.07**	**0.09**	**0.13**	0.09	**0.07**	**0.09**	**0.07**	**0.15**	**0.30**	**0.29**	**0.26**	0.00		0.00	**0.08**
COC	**0.14**	**0.16**	**0.17**	**0.15**	**0.10**	**0.11**	**0.12**	**0.09**	**0.09**	**0.11**	**0.10**	**0.14**	**0.28**	**0.27**	**0.26**	0.00	0.00		**0.07**
CLIP	**0.15**	**0.17**	**0.22**	**0.18**	**0.12**	**0.18**	**0.15**	**0.27**	**0.10**	**0.12**	**0.11**	**0.16**	**0.30**	**0.32**	**0.29**	0.03	**0.09**	**0.07**	

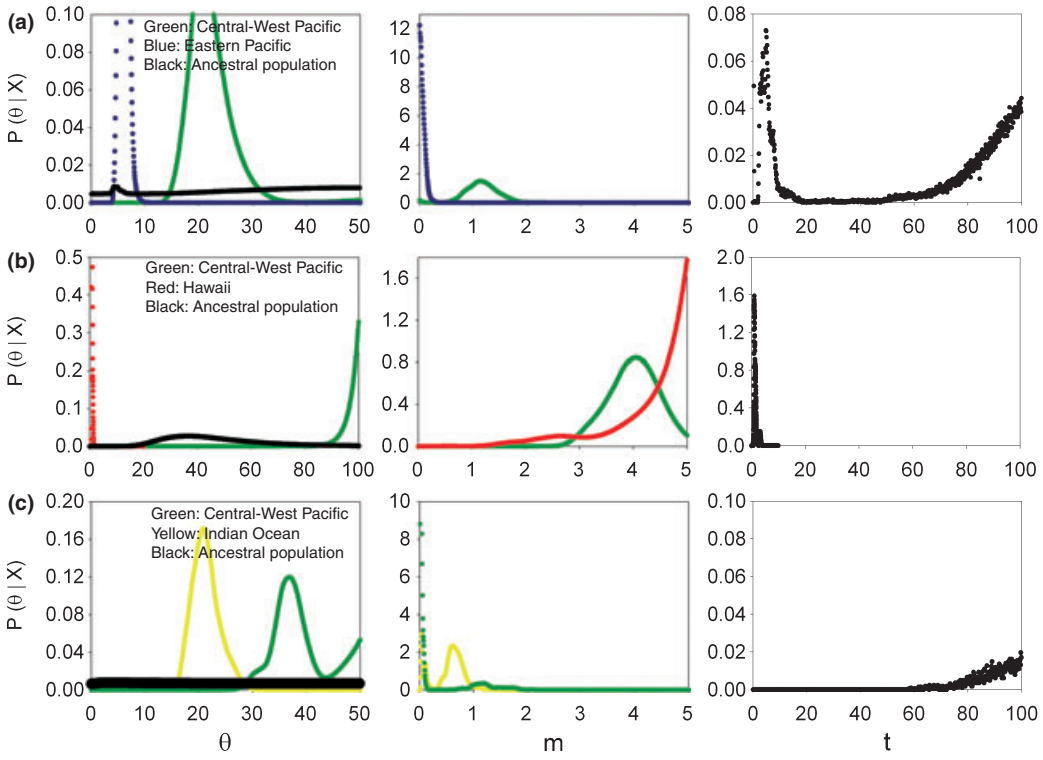

Fig. 4 The posterior probabilities of parameter estimates from three IMa models. (a) Central-West Pacific and eastern Pacific. (b) Central-West Pacific and Hawaii. (c) Central-West Pacific and Indian Ocean.

Other studies of similar geographical scope have focused on using mitochondrial DNA to delineate morphologically cryptic species complexes that were formally thought to represent broadly distributed species (e.g. Bowen *et al*. 2001; Colborn *et al*. 2001). Our data allow us to examine contrasting geographical patterns of population divergence and migration dynamics in relation to the two opposing hypotheses for the origins of the IAA diversity hotspot: the Center of Origin and the Center of Accumulation models. Contrary to the idea that the IAA acts as a centre of origin our multilocus data indicates that, in *Scarus rubroviolaceus*, two major recent divergence events have occurred at isolated peripheral areas: the Hawaiian Islands and the tropical eastern Pacific. Although we detected population division between the Pacific and Indian Oceans predicted by centre of origin model, the signal of this splitting event is much weaker than the in the first two cases. Second, the IMa model reveals a pattern of predominantly westerly migration from the eastern Pacific to the Central Pacific, a major assumption of the Center of Accumulation model. In the tropical Pacific, northern

Table 6 Maximum-likelihood estimates (MLE) of isolation and migration model parameters and their respective demographic conversions for the Central-West Pacific (population 1) and eastern Pacific (population 2). The demographic estimates are as follows: N_1, effective population size of the Central-West Pacific; N_2, effective population size of the eastern Pacific; $2N_1m_1$, number of migrants per generation migrating into the Central-West Pacific from the eastern Pacific; $2N_2m_2$, number of migrants per generation migrating into the eastern Pacific from the Central-West Pacific. The 90% high point density (HPD) is the shortest span along the axes of Fig. 4A that contain 90% of the histograms. Asterisks for HPD estimates indicate cases where the probability distributions for priors do not include 0.0. Demographic conversions are based on a geometric mean mutation rate of 0.00028/year and generation time of 5 years (Howard 2008). See text and Appendix S5 for details

	Model parameters						Demographic conversions					
Estimate	θ_1	θ_2	θ_A	m_1	m_2	t	N_1	N_2	N_A	$2N_1m_1$	$2N_2m_2$	Splitting time (years)
MLE	20.43	5.88	4.275	1.133	0.003	4.850	3681	1059	770	11.57	0.007	17 483
Lower 90% HPD	16.18	4.78	3.68*	0.723	0.003	0.25*	2915	861				
Upper 90% HPD	27.73	7.23	148.88*	1.623	0.128	99.95*	4997	1302				

Table 7 Tests of nested models for the Central Pacific and eastern Pacific. Shaded tests do not improve fit compared to the full model

Model (Θ)	$Log(p'(\hat{\Theta}\|X))$	$-2\hat{\Lambda}$*	P	d.f.
Full model	−2.616			
$m_1 = m_2$	−25.176	45.122	<0.001	1
$m_2 = 0$	−2.494	0.244	0.621	1†
$m_1 = 0$	2.285	9.801	0.002	1†
$m_1 = 0$, $m_2 = 0$	2.284	9.799	0.007	2†
$\theta_1 = \theta_2$	−227.135	449.038	<0.001	1
$\theta_1 = \theta_2 = \theta_A$	−227.168	449.105	<0.001	2
$\theta_1 = \theta_2$; $m_1 = m_2$	−460.517	915.802	<0.001	2
$\theta_1 = \theta_2$; $m_1 = 0$; $m_2 = 0$	−460.517	915.802	<0.001	3†
$\theta_1 = \theta_2 = \theta_A$; $m_1 = m_2$	−460.517	915.802	<0.001	3
$\theta_1 = \theta_2 = \theta_A$; $m_1 = 0$; $m_2 = 0$	−460.517	915.802	<0.001	4†
$\theta_1 = \theta_A$	−3.458	1.682	0.195	1
$\theta_1 = \theta_A$; $m_1 = m_2$	−25.142	45.052	<0.001	2
$\theta_1 = \theta_A$; $m_1 = 0$; $m_2 = 0$	−460.517	915.802	<0.001	3†
$\theta_2 = \theta_A$	−3.456	1.684	0.194	1
$\theta_2 = \theta_A$; $m_1 = = m_2$	−25.176	45.112	<0.001	2
$\theta_2 = \theta_A$; $m_1 = 0$; $m_2 = 0$	−26.582	47.934	<0.001	3†

*Twice the log likelihood ratio is distributed as χ^2.
†Test distribution of $-2\hat{\Lambda}$ is a mixture.

and southern gyres that rotate in opposite directions produce strong westward flow that could facilitate westward larval dispersal along the equator towards the IAA at the western edge of the Pacific. Lessios & Robertson (2006) found a predominance of westward gene flow from the eastern Pacific in a variety of reef fishes whose eastern Pacific populations originated in the Central Pacific and proposed that that bias is owing to the existence of larger recipient populations in the Central Pacific. Our data on the directional bias in gene flow and much larger population size of *S. rubroviolaceus* in the Central Pacific are consistent with that hypothesis. The combined effects of peripheral diversification and subsequence migration towards the IAA would be to increase total genetic diversity along the western margin of the Pacific Ocean. In theory at least, this diversification and migration dynamic could also lead to a species accumulation in the IAA. However, if such a process made a major contribution to the gradient in diversity across the Pacific, there would have to have been higher species turnover in peripheral areas. As yet there is no evidence for or against such a process.

As in all models, departures from assumptions may bias parameter estimates. We can confidently show that none of the 15 loci are tightly linked, and we could control for departures from neutrality by constructing sets of loci that did not show evidence for strong selection in BayeScans. Regarding the SMM assumed by IMa, we rejected the SMM in favour of a more complex TPM for only three loci. Two of these made it past BayeScans and into IMa data sets. However, the overall effect of these loci in terms of the fraction of alleles generated by nonstepwise mutations was on the order of a few per cent. Thus, we can conclude that a SMM was a good fit to the microsatellite loci used in the IMa models. A second IMa assumption warrants closer examination that un-sampled populations are not acting as alternative routes for migrants (Hey & Nielsen 2004). In the Central-West Pacific vs. eastern Pacific comparison, the only candidate for such a bridging population is Hawaii, but the large degree of differentiation between Hawaii and all other populations ($F_{ST} > 0.10$, Table 3) makes it unlikely that a Hawaii 'ghost population' is receiving gene copies from the eastern Pacific and then exporting them to the broader Central-West Pacific at any significant rate. Recent simulations by Strasburg & Rieseberg (2010) have shown that even when low levels of gene flow occur via a third ghost population ($N_em < 0.2$) biases in the migration rates between the two focal populations are small. Summarizing, small departures from SMM and the existence of ghost populations are unlikely to strongly bias the model parameters that we could estimate.

The >5000 km of deep water that divides the eastern Pacific from the nearest Islands of the Central Pacific is potentially one of the greatest challenges to larval dispersal in the oceans and is known as the eastern Pacific Barrier or EPB (Ekman 1953). Our IMa migration parameters for *S. rubroviolaceus* indicate high gene flow across this 'impassable' barrier in a westward direction, but essentially no gene flow in the opposite direction. Lessios & Robertson (2006) also fitted an IMa model to a mtDNA data set across the EPB in *S. rubroviolaceus*, which placed the most recent split at ~30 000 BP, but found no clear asymmetry in the direction of gene flow. Given the effects of even slight errors in estimating an average mutation rate on demographic conversions, our estimate (t ~17 000 years) is comparable. These differences in directionality of geneflow estimates are probably related to the fact that Lessios and Robertson's used data from two mitochondrial ATPase genes, which rarely recombine. Our conclusion of asymmetric migration across the EPB in *S. rubroviolaceus* is based on the random distribution of coalescent events provided by multiple unlinked nuclear loci. More comparative, multilocus studies are needed before generalizations can be made about the timing, magnitude and direction of dispersal across major potential oceanic barriers to marine dispersal.

Our analyses indicate that the genetic discontinuity or 'break' between the Pacific and Indian Oceans at the IAA has a weak effect on the data compared to the isolation of Hawaii or the eastern Pacific. A suite of

studies, working on a broad diversity of marine organisms, has found an absence of genetic structure between these oceans (see reviews in Williams *et al.* 2002; Crandall *et al.* 2008). There are two general explanations for why some species show weak or no structure across this potential historical barrier to gene flow: (i) species with high dispersal potential maintained high connectivity across narrow channels that probably spanned the IAA region during low sea level stands; or (ii) species with high dispersal potential expanded their populations rapidly into other areas after subdivision caused by low sea levels. However, comparative studies of closely related species with similar potentials for larval dispersal can show markedly different patterns of genetic structure between the Indian and Pacific Oceans. For example, in congeneric sea urchins (Lessios *et al.* 2001) and sea snails (Crandall *et al.* 2008), one species will have a clear genetic break between the Indian Ocean and the West Pacific, while another will not. Subtle and undetected differences in larval biology or slight differences in adult habitat may determine which species moves freely between ocean basins (Crandall *et al.* 2008). In this light, it is instructive to compare our results with those from the only other large-scale, population genetic study of a parrotfish. *Chlorurus sordidus* is a reef-associated parrotfish that occurs in the Indian and Pacific Oceans (including Hawaii) but not in the eastern Pacific. The adult morphology and ecology of *C. sordidus* and *S. rubroviolaceus* are quite similar (Bellwood & Choat 1990), and they have similar larval durations (Chen 1999; B. Victor personal communication). Bay *et al.* (2004) found two major divergent populations within *C. sordidus*: one in the Western Indian Ocean and a second in the Eastern Indian Ocean plus the remainder of the Pacific. Within the latter clade, there was evidence of only minor divergence between Hawaii and the rest of the Central-West Pacific. This pattern of structure in *C. sordidus* is essentially the inverse of that we find in *S. rubroviolaceus* and indicates that only *C. sordidus* responded to some putative vicariance event between the Indian and Pacific basins by differentiating to the point of no gene flow between clades in contemporary populations.

There is another feature in our genetic data that raises additional questions about the mechanism(s) that result in diversification between the Pacific and Indian Oceans. In *C. sordidus*, the geographical break is at the eastern edge of the Indian Ocean, between Western Australia and Cocos-Keeling. In *S. rubroviolaceus*, however, the break occurs further west: most of our samples from the Cocos-Keeling Islands and Christmas Island were assigned to the population that includes the Central and West Pacific Ocean, rather than the population defined by samples from western Thailand and the Western Indian Ocean. Some marine invertebrates with broad Indo-Pacific distributions also have populations along Northern and Western Australia that are panmictic with the West Pacific rather than the West Indian Ocean (Lessios *et al.* 2001; Williams *et al.* 2002). The typical explanation for this pattern is that during transgressions, the Indonesian thoroughflow disperses larvae westward across the IAA where they are entrained in the warm Leeuwin current which travels south along coastal Western Australia. The occurrence of a substantial number of *S. rubroviolaceus* that are clearly assigned to the Central-West Pacific at Cocos-Keeling and Christmas Island, as well as a recently discovered hybrid zone between sister species of reef fish at this same location (Hobbs *et al.* 2009), indicates that either dispersal has occurred much deeper into the Indian Ocean during high periods of high sea level across the IAA, or the mechanism of population splitting and speciation is not related to the coastal environments of the IAA. More comparative studies of sets of closely related species that involve synoptic sampling of the complex and extended eastern boundary of the Indian Ocean could help resolve that issue.

Species-level phylogenies of various invertebrates point to peripheral areas beyond the IAA as important centres of origination of diversity through peripatric speciation (*Echinolittorina*: Williams & Reid 2004; *Nerita*: Frey & Vermeij 2008; *Calcinus*: Malay & Paulay 2009). Our study provides clear population-level support for this mode of speciation. A caveat to our approach is that we cannot be certain that these remote populations in Hawaii and the eastern Pacific will evolve reproductive isolation from their sister group. However, we can deduce from the biogeographical pattern of high levels of endemism in both regions (Robertson & Allen 2008; Randall 1998) that the end result of such extreme geographical isolation can be new species. As a potential window into the early phases of speciation, the Hawaii and eastern Pacific *S. rubroviolaceus* populations could be used to understand the molecular and phenotypic changes associated with the earliest phases of speciation. Sexual selection has been hypothesized to be rapid driver of speciation parrotfish that exhibit sexual dichromatism and pronounced colour variation between closely related species (Kazancioglu *et al.* 2009). The level of phenotypic divergence associated with molecular divergence is unknown, but this and other parrotfish systems provide excellent opportunities to use tools from the emerging field of population genomics to understand the roles of drift and selection in early population divergence and to link molecular evolution to subtle potential phenotypic change (Nielsen *et al.* 2009). From a more comprehensive population-level understanding of this relatively short, but exceedingly

important 2.5 Ma period of evolutionary history (Hewitt 2000), we stand to gain a richer mechanistic understanding of the complex processes that lead to new species and drive diversity gradients in the tropics.

Acknowledgements

We thank the following organizations for facilitating field work and help with permits: National Research Council of Thailand; Department of Marine and Wildlife Resources-American Samoa; Department of Agriculture, Forests, and Fisheries-Western Samoa; and the Hawaii Department of Aquatic Resources (DAR). We thank B. Victor for providing us with 50 samples from Cocos Island in the eastern Pacific; G. Matsuda of the Alii Holo Kai Dive Club for allowing us to sample at spear fishing tournaments on Hawaii; and S. Hau and T. Beirne for additional samples from Maui and the Big Island respectively. We thank B. Bowen, J. H. Choat, B. Holland, D. Kapan, D. Rubinoff, and three anonymous reviewers for comments on the manuscript. This study was partially supported by the National Oceanic and Atmospheric Administration, Center for Sponsored Coastal Ocean Research, under awards #NAO06NOS4260200 to the University of Hawaii for the Hawaii Coral Reef Initiative. Fieldwork in Samoa and Thailand was supported by a grant from the Ecology, Evolution and Conservation Biology Program at the University of Hawaii (EECB) awarded to JMF. The EECB is funded by a NSF GK-12 program grant #NSF DGE02-32016 to K.Y. Kaneshiro. Collecting activity by DRR was supported by National Geographic Society grants NGS 5831-96 and 7269-02.

References

Alfaro ME, Brock CD, Banbury BL, Wainwright PC (2009) Does evolutionary innovation in pharyngeal jaws lead to rapid lineage diversification in labrid fishes? *BMC Evolutionary Biology*, **9**, 255.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates, Sunderland, MA.

Barber PH, Erdmann MV, Palumbi SR (2006) Comparative phylogeography of three codistributed stomatopods: origins and timing of regional lineage diversification in the coral triangle. *Evolution*, **60**, 1825–1839.

Bay LK, Choat JH, van Herwerden L, Robertson DR (2004) High genetic diversities and complex genetic structure in an Indo-Pacific tropical reef fish (*Chlorurus sordidus*): evidence of an unstable evolutionary past? *Marine Biology*, **144**, 757–767.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**, 251–261.

Bellwood DR, Choat JH (1990) A functional analysis of grazing in parrot fishes (family Scaridae): the ecological implications. *Environmental Biology of Fishes*, **28**, 189–214.

Benzie JAH (1999) Genetic structure of coral reef organisms: ghosts of dispersal past. *American Zoologist*, **39**, 131–145.

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR (2001) Phylogeography of the trumpetfishes (Aulostomus): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Briggs JC (1999) Coincident biogeographic patterns: Indo-West Pacific Ocean. *Evolution*, **53**, 326–335.

Briggs JC (2003) Marine centres of origin as evolutionary engines. *Journal of Biogeography*, **30**, 1–18.

Carlon DB, Lippe C (2007) Isolation and characterization of 17 new microsatellite markers for the ember parrotfish *Scarus rubroviolaceus*, and cross-amplification in four other parrotfish species. *Molecular Ecology Notes*, **7**, 613–616.

Chen LS (1999) *Ontogenetic development in post settlement scarids (Pisces: Scaridae)*. PhD Dissertation, James Cook University, Townsville, Australia.

Colborn J, Crabtree RE, Shaklee JB, Pfeiler E, Bowen BW (2001) The evolutionary enigma of bonefishes (*Albula* spp.): cryptic species and ancient separations in a globally distributed shorefish. *Evolution*, **55**, 807–820.

Crandall ED, Frey MA, Grosberg RK, Barber PH (2008) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology*, **17**, 611–626.

Dieringer D, Schlotterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Dyer RJ (2009) Genetic Studio: a suite of programs for spatial analysis of genetic-marker data. *Molecular Ecology Resources*, **9**, 110–113.

Dyer RJ, Nason JD (2004) Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1727.

Ekman S (1953) *Zoogeography of the Sea*. Sidgwick and Jackson, London.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.

Estoup A, Cornuet J-M (1999) Microsatellites: evolution and applications. In: *Microsatellite Evolution: Inferences from Population Data* (eds Goldstein DB, Schlotterer C), pp. 49–64. Oxford University Press, Oxford.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977.

Frey MA, Vermeij GJ (2008) Molecular phylogenies and historical biogeography of a circumtropical group of gastropods (Genus: *Nerita*): implications for regional diversity patterns in the marine tropics. *Molecular Phylogenetics and Evolution*, **48**, 1067–1086.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D-persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.

Hobbs JPA, Frisch AJ, Allen GR, Van Herwerden L (2009) Marine hybrid hotspot at Indo-Pacific biogeographic border. *Biology Letters*, **5**, 258–261.

Howard KG (2008) *Community structure, life history, and movement patterns of parrotfishes: large protogynous fishery species*. PhD Dissertation, University of Hawaii, Honolulu, HI, USA.

Hughes TP, Bellwood DR, Connolly SR (2002) Biodiversity hotspots, centres of endemicity, and the conservation of coral reefs. *Ecology Letters*, **5**, 775–784.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801.

Jokiel P, Martinelli FJ (1992) The vortex model of coral reef biogeography. *Journal of Biogeography*, **19**, 449–458.

Jones AG, Rosenqvist G, Berglund A, Avise JC (1999) Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics*, **152**, 1057.

Kazancioglu E, Near TJ, Hanel R, Wainwright PC (2009) Influence of sexual selection and feeding functional morphology on diversification rate of parrotfishes (Scaridae). *Proceedings Biological Sciences/The Royal Society*, **276**, 3439–3446.

Lambeck K, Esat TM, Potter EK (2002) Links between climate and sea levels for the past three million years. *Nature*, **419**, 199–206.

Leopoldino AM, Pena SDJ (2003) The mutational spectrum of human autosomal tetranucleotide microsatellites. *Human Mutation*, **21**, 71–79.

Lessios HA, Robertson DR (2006) Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings. Biological Sciences/The Royal Society*, **273**, 2201–2208.

Lessios HA, Kessing BD, Pearse JS (2001) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution*, **55**, 955–975.

Lessios HA, Kane J, Robertson DR (2003) Phylogeography of the pantropical sea urchin *Tripneustes*: contrasting patterns of population structure between oceans. *Evolution*, **57**, 2026–2036.

Losos JB, Glor RE (2003) Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution*, **18**, 220–227.

Malay MCMD, Paulay G (2009) Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: *Calcinus*). *Evolution*, **64**, 632–664.

Meyer CP, Geller JB, Paulay G (2005) Fine scale endemism on coral reefs: archipelagic differentiation in turbinid gastropods. *Evolution*, **59**, 113–125.

Mora C, Chittaro PM, Sale PF, Kritzer JP, Ludsin SA (2003) Patterns and processes in reef fish diversity. *Nature*, **421**, 933–936.

Nielsen R (1997) A likelihood approach to populations samples of microsatellite alleles. *Genetics*, **146**, 711–716.

Nielsen R, Palsbøll PJ (1999) Single-locus tests of microsatellite evolution: multi-step mutations and constraints on allele size. *Molecular Phylogenetics and Evolution*, **11**, 477–484.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology*, **18**, 3128–3150.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Parenti P, Randall JE (2000) An annotated checklist of the species of the labroid fish families Labridae and Scaridae. *Ichthyological Bulletin*, **68**, 1–97.

Pavey SA, Nielsen JL, Hamon TR (2009) Recent ecological divergence despite migration in sockeye salmon (*Oncorhynchus nerka*). *Evolution*, **64**, 1773–1783.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Randall JE (1998) *Shore Fishes of Hawaii*. University of Hawaii Press, Honolulu, HI.

Robertson DR, Allen GR (2008) Shorefishes of the tropical Eastern Pacific online information system. http://www.stri.org/sftep

Robertson DR, Grove JS, McCosker JE (2004) Tropical transpacific shore fishes. *Pacific Science*, **58**, 507–565.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Steinberg EK, Lindner KR, Gallea J, Maxwell A, Meng J, Allendorf FW (2002) Rates and patterns of microsatellite mutations in pink salmon. *Molecular Biology and Evolution*, **19**, 1198.

Storey J, Storey JD, Tibshirani R (2003) Statistical significance for genome-wide experiments. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440–9445.

Strasburg JL, Rieseberg LH (2010) How robust are''Isolation with Migration'' analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution*, **27**, 297–310.

Streelman JT, Alfaro M, Westneat MW, Bellwood DR, Karl SA (2002) Evolutionary history of the parrotfishes: biogeography, ecomorphology, and comparative diversity. *Evolution*, **56**, 961–971.

Williams ST, Duda TF (2008) Did tectonic activity stimulate Oligo-Miocene speciation in the Indo-West Pacific? *Evolution*, **62**, 1618–1634.

Williams ST, Reid DG (2004) Speciation and diversity on tropical rocky shores: a global phylogeny of snails of the genus *Echinolittorina*. *Evolution*, **58**, 2227–2251.

Williams ST, Jara J, Gomez E, Knowlton N (2002) The marine Indo-West Pacific break: contrasting the resolving power of mitochondrial and nuclear genes. *Integrative and Comparative Biology*, **42**, 941–952.

Woodland DJ (1983) Zoogeography of the Siganidae (Pisces): an interpretation of distribution and richness patterns. *Bulletin of Marine Science*, **33**, 713–717.

This paper was part of the MS work of JMF, who is interested in the evolutionary ecology of pelagic and near-shore fisheries and conservation of marine environments in his native Hawaii. DC, supervised this work and is broadly interested in how new species form in the species-rich tropics. CL shares interests in molecular ecological approaches to conservation of freshwater and marine fish. DRR's research spans the ecology and evolution of reef fishes.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Population statistics for all collecting localities with $n > 5$ (excludes Oman, Okinawa, and Rota Island). *N*, number of samples, *Ng*, number of samples genotyped for each locus (na, not attempted because of insufficient quantity of DNA template for PCR), *A*, number of alleles, H_{obs}, observed heterozygosity, H_{exp} expected heterozygosity, F_{IS}, Wrights inbreeding coefficient. Probability values are for the null hypothesis that $F_{IS} > 0.0$. Cells with 'np' indicate no polymorphism for the estimate/test. The prefix 'Sru' is not included with locus names.

Appendix S2 The frequencies of null alleles at two loci in four populations. Sample groupings based on STRUCTURE, $K = 4$ model. Frequencies were calculated with the software MICRO-CHECKER based on the estimator of van Oosterhout *et al.* (2004).

Appendix S3 Pairs of loci within sampling localities with significant linkage disequilibrium (LD) as indicated by *Q* values <0.01. LD. Shading indicates pairs of loci that were significant in both the Cocos (Keeling) sample and the Seychelles.

Appendix S4 Maximum-likelihood estimates of IMa parameters from replicate runs of three IMa models. The 90% high point density (HPD) is the shortest span along the axis of the posterior distribution that contains 90% of the histograms. Asterisks for HPD estimates indicate cases where the posterior probability distributions do not include 0. Because of the computational limitations, there was only one run of the Central-West Pacific vs. Indian Ocean comparison. The # of trees column is the number of gene genealogies sampled over the run. For replicate runs, both were used in the L-mode of IMa to generate joint posterior probabilities presented in Table 6.

Appendix S5 Estimation of the mutation rate using scalars from the IMa model of the Central-West Pacific and eastern Pacific. The mutation rate (μ) for Locus *B115* was set to 0.00094 per generation as measured by Jones *et al.* (1999) in a pedigree analysis of a pipefish tetra-nucleotide locus of similar length. The generation time of *Scarus rubroviolaceus* is estimated to be 5 years (Howard 2008). See text for details.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.