

VOLUME 24
NUMBER 7
APRIL
2015

MOLECULAR ECOLOGY



Genomic signatures of geographic isolation and natural selection in coral reef fishes

MICHELLE R. GAITHER,*† MOISÉS A. BERNAL,†‡ RICHARD R. COLEMAN,§ BRIAN W. BOWEN,§ SHELLEY A. JONES,† W. BRIAN SIMISON¶ and LUIZ A. ROCHA†

*School of Biological and Biomedical Sciences, Durham University, South Road, Durham DH1 3LE, UK, †Section of Ichthyology, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, CA 94118, USA, ‡University of Texas Marine Science Institute, 750 Channel View Drive, Port Aransas, TX 78373, USA, §Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa, P.O. Box 1346, Kane'ohe, HI 96744, USA, ¶Center for Comparative Genomics, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, CA 94118, USA

Abstract

The drivers of speciation remain among the most controversial topics in evolutionary biology. Initially, Darwin emphasized natural selection as a primary mechanism of speciation, but the architects of the modern synthesis largely abandoned that view in favour of divergence by geographic isolation. The balance between selection and isolation is still at the forefront of the evolutionary debate, especially for the world's tropical oceans where biodiversity is high, but isolating barriers are few. Here, we identify the drivers of speciation in Pacific reef fishes of the genus *Acanthurus* by comparative genome scans of two peripheral populations that split from a large Central-West Pacific lineage at roughly the same time. Mitochondrial sequences indicate that populations in the Hawaiian Archipelago and the Marquesas Islands became isolated approximately 0.5 Ma. The Hawaiian lineage is morphologically indistinguishable from the widespread Pacific form, but the Marquesan form is recognized as a distinct species that occupies an unusual tropical ecosystem characterized by upwelling, turbidity, temperature fluctuations, algal blooms and little coral cover. An analysis of 3737 SNPs reveals a strong signal of selection at the Marquesas, with 59 loci under disruptive selection including an opsin Rh2 locus. While both the Hawaiian and Marquesan populations indicate signals of drift, the former shows a weak signal of selection that is comparable with populations in the Central-West Pacific. This contrast between closely related lineages reveals one population diverging due primarily to geographic isolation and genetic drift, and the other achieving taxonomic species status under the influence of selection.

Keywords: adaptation, genetic drift, natural selection, outlier loci, RADSeq, speciation

Received 10 August 2014; revision received 20 February 2015; accepted 24 February 2015

Introduction

Speciation, or the evolutionary process by which new species arise, continues to be one of the most controversial topics in evolutionary biology. Views on how species are formed have changed dramatically in the last 150 years. Darwin's idea that new species form by disruptive selection was heavily criticized by the architects

of the modern synthesis and almost abandoned in view of species originating by geographic isolation (Darwin 1859; Coyne & Orr 2004). However, with the support of numerous examples, ecological speciation is now widely accepted (Schluter 2009; Nosil 2012; Bowen *et al.* 2013) but the relative importance of geographic isolation, and the genetic changes that lead to speciation, is still hotly debated topic in evolutionary biology (Nosil *et al.* 2009; Via 2009; Supple *et al.* 2014). With recent access to genomic information, the traditional view of a cohesive, gradually evolving genome is giving way to a

Correspondence: Michelle R. Gaither, Fax: +44 01913341323; E-mail: michellergaither@gmail.com

view of permeable genomes, wherein islands of genomic differentiation are linked to loci under disruptive selection (Wu 2001; Via 2012). This pattern is usually associated with sympatric speciation (Michel *et al.* 2010); however, ecological speciation is not tied to a specific geographical context (Rundle & Nosil 2005). Furthermore, natural selection can drive populations to reproductive isolation with or without gene flow (Schluter 2001; Nosil 2012).

Marine organisms are a particular challenge to models of speciation that require isolation for extended periods of time as many of these taxa have large population sizes, vast geographic ranges and long pelagic larval phases that extend for weeks to months, potentially connecting distant habitats (Bowen *et al.* 2013). Furthermore, there is a relative paucity of strong barriers to dispersal in the oceans, with recognized barriers often being porous in nature and sporadically traversed when conditions are favourable (Briggs & Bowen 2013; Gaither & Rocha 2013). These observations, combined with the high biodiversity on coral reefs, prompt the conclusion that natural selection, perhaps reinforced by assortative mating (e.g. Puebla *et al.* 2014), must play an important role in speciation in the sea.

A prerequisite for assortative mating is mate recognition, and among the often brilliantly coloured fishes found on coral reefs, visual cues are thought to play a key role (Helfman *et al.* 2009). While colour patterns are expected to diverge slowly under genetic drift, they can evolve rapidly when this character is involved in mate choice (Endler & Houde 1995; McMillan *et al.* 1999; Puebla *et al.* 2007, 2014). In the Caribbean reef fishes known as hamlets (genus *Hypoplectrus*), distinct colour morphs are thought to be maintained by aggressive mimicry and cryptosis, reinforced by assortative mating (Puebla *et al.* 2011, 2014), whereas in African cichlids (family Cichlidae), sexual selection on male nuptial coloration plays a role in isolating populations (Salzburger *et al.* 2006; Blais *et al.* 2009; Selz *et al.* 2014). In both cases, shifts in colour patterns driven by natural selection outpace genetic divergence at neutral loci; however, there are also abundant counter-examples (McCartney *et al.* 2003; Rocha 2004; Messmer *et al.* 2005; Craig *et al.* 2006; Drew & Barber 2009; Leray *et al.* 2010; DiBattista *et al.* 2012).

With high species richness and recent evolutionary radiations (Choat *et al.* 2012; Sorenson *et al.* 2013; Gaither *et al.* 2014), tropical oceans are an excellent setting for the study of ecological speciation, yet there are only a few case studies (Munday *et al.* 2004; Rocha *et al.* 2005; Johannesson 2009; Bird *et al.* 2011; Prada & Hellberg 2013). Evidence for selection has been inferred from shifts in habitat use (Munday *et al.* 2004; Bird *et al.* 2011), differences in survivorship (Prada & Hellberg 2013) or divergences at neutral loci that correspond

with ecological rather than geographic partitions (Rocha 2004). Few studies have yet provided the genomic evidence to support these inferences, but initial results are promising. In hamlets, genome scans indicate that *Hox* gene evolution may be involved in the radiation of distinct morphotypes (Puebla *et al.* 2014). In freshwater fishes, compelling cases of divergence driven by natural selection are found in three-spine sticklebacks (*Gasterosteus aculeatus*), where genomic islands of divergence correspond to colonization of novel habitats (Hohenlohe *et al.* 2010) and in lake whitefish (*Coregonus clupeaformis*), where divergence coincides with niche partitioning (Renaut *et al.* 2012). Few studies, however, have explored the contrast between the genomic signatures of differentiation caused by geographic isolation and disruptive selection in marine organisms.

To resolve the genomic signatures of isolation and selection, we analysed two members of the Orangeband Surgeonfish species complex: *Acanthurus olivaceus* Bloch and Schneider 1801 and *A. reversus* Randall and Earl 1999. The widely distributed *A. olivaceus* occurs across the central and western Pacific and extends to the Hawaiian Archipelago, while *A. reversus*, recently described as a separate species, is endemic to the Marquesas (Fig. 1, Randall & Earle 1999). Only slight differences in fin ray counts distinguish *A. reversus* from *A. olivaceus* (Table S1, Supporting information), but they can easily be differentiated based on colour patterns (Randall 2005; Fig. 1a). These species, like most surgeonfishes, inhabit shallow waters around coral reefs (Robertson 1983). It is estimated that *A. olivaceus* can live for at least 35 years (Choat & Axe 1996), yet as in most coral reef organisms dispersal is limited to a pelagic larval stage (Leis 1991; Leis & McCormick 2002). The pelagic larval duration for our study species has not been determined but estimates for other Acanthurids range from 40 to 70 days (Doherty *et al.* 1995; Fisher *et al.* 2005).

Initially a targeted locus approach, including mtDNA cytochrome *b* (*Cytb*), and nuclear introns *S7* and *Gpd2* (data reported here), revealed that the isolated Hawaiian population of *A. olivaceus* began diverging from the Central-West Pacific populations at roughly the same time as the Marquesan *A. reversus*: a surprising result given the taxonomic status and unique coloration of *A. reversus* (Fig. 1a). In light of these findings, we analysed 3737 single nucleotide polymorphisms (SNPs) obtained through RADSeq (Restriction Site Associated DNA Sequencing) (Davey *et al.* 2011; Peterson *et al.* 2012). These genome-wide surveys were used to detect signals of selection and isolation in the Hawaiian and Marquesan lineages, to answer two questions: Is it possible to distinguish between isolation and selection in these recently diverged species? Has selection accelerated evolutionary divergence at the Marquesas?

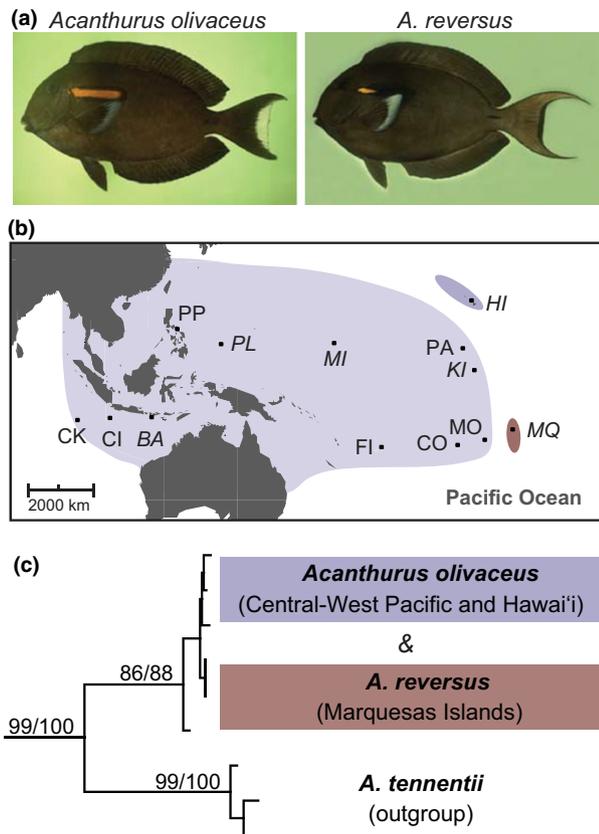


Fig. 1 (a) Coloration of Central-West Pacific and Hawaiian populations of *Acanthurus olivaceus* versus the Marquesan endemic *A. reversus*. (b) Map of study area and geographic range of each species (blue, *A. olivaceus*; red, *A. reversus*). Collection locations include Hawai'i (HI), Palmyra (PA), Kiritimati (KI), Moorea (MO), Cook Islands (CO), Fiji (FI), Marshall Islands (MI), Palau (PL), Philippines (PP), Bali (BA), Christmas Island (CI), Cocos-Keeling (CK), and Marquesas Islands (MQ). Locations in italics were sub-sampled and amplified at 3737 RADSeq loci. (c) Maximum-likelihood tree generated using cytochrome *b* sequences and rooted using *A. tennentii*. Bootstrap support values and Bayesian posterior probabilities are presented (see Fig. S1, Supporting information for complete tree). Photo credit: J.E. Randall.

Materials and methods

Specimen collection

A total of 302 specimens of *A. olivaceus* and 30 specimens of *A. reversus* were collected either by SCUBA divers using pole spears or obtained from local fish markets where the provenance of the fish was known (Table 1). Gill, fin and/or muscle tissues were preserved in either salt-saturated DMSO or 95% ethanol and stored at room temperature at the Hawai'i Institute of Marine Biology.

Sanger sequence analyses

Laboratory protocols for Sanger sequencing are provided in Supplemental information. Summary statistics for the mtDNA, including haplotype diversity (h) and nucleotide diversity (π), were estimated using algorithms from Nei (1987) as implemented in the software package ARLEQUIN 3.5 (Excoffier *et al.* 2005). The AIC implemented in jMODELTEST 0.1.1 (Posada 2008) indicated the TPM3uf+G as the best fit model of DNA sequence evolution with a gamma value of 0.32. Median-joining networks were constructed using the program NETWORK 4.5 with default settings (Bandelt *et al.* 1999). To evaluate the phylogenetic relationship between the study species, we used the maximum-likelihood (ML) method and default settings of the program MEGA 5.0 (Tamura *et al.* 2011). *Cytb* sequences for surgeonfishes of the genus *Acanthurus* (Acanthuridae) were obtained from GenBank (Table S2, Supporting information) including the closely related *A. tennentii* (Sorenson *et al.* 2013), and trees were rooted using sequence data from the Spotted Surgeonfish, *Ctenochaetus strigosus* and the Elongate Unicornfish, *Naso lopezi*. Bootstrap support values were calculated using default settings with 2000 replicates. The ML tree topology was confirmed by Bayesian Markov Chain Monte Carlo (MCMC) analysis using MRBAYES 3.1.1 (Ronquist & Huelsenbeck 2003). The Bayesian analysis was run using the GTR model with gamma-distributed rate variation across sites for two million iterations sampling every 100th generation with an initial burn-in of 5000 samples. We calculated the divergence between mitochondrial lineages (d) in ARLEQUIN.

To test for genetic structure within and between species, an analysis of molecular variance (AMOVA) was performed in ARLEQUIN using 20 000 permutations. Because the TPM3uf+G model of sequence evolution is not implemented in ARLEQUIN, we used the most similar model available (TrN; Tamura & Nei 1993) with a gamma value of 0.32. An analogue of Wright's F_{ST} , Φ_{ST} , which incorporates a model of sequence evolution and distance, was calculated for the entire data set and for pairwise comparisons among all locations. We maintained $\alpha = 0.05$ among all pairwise tests by controlling for the false discovery rate (Narum 2006).

To estimate the time to coalescence based on the mtDNA data set, we used the Bayesian MCMC approach implemented in BEAST 1.7.4 (Drummond & Rambaut 2007). We conducted our analysis with a relaxed lognormal clock and uncorrelated substitution rates among branches. We used default priors under the HKY + G model of mutation and ran simulations for 10 million generations with sampling every 1000

Table 1 Molecular diversity indices for populations of *Acanthurus olivaceus* and *A. reversus*

Sample location	<i>Cytb</i>				<i>S7</i>					<i>Gpd2</i>				
	<i>N</i>	<i>N_h</i>	<i>h</i>	π	<i>N</i>	<i>N_a</i>	<i>H_O</i>	<i>H_E</i>	<i>P</i> -value	<i>N</i>	<i>N_a</i>	<i>H_O</i>	<i>H_E</i>	<i>P</i> -value
<i>Acanthurus olivaceus</i>														
Hawai'i (HI)	35	3	0.51 ± 0.05	0.003 ± 0.002	32	6	0.69	0.62	0.958	35	6	0.57	0.65	0.358
Palmyra (PA)	26	11	0.75 ± 0.09	0.004 ± 0.003	30	7	0.73	0.71	0.735	29	4	0.69	0.68	0.314
Kiritimati (KI)	32	16	0.86 ± 0.05	0.005 ± 0.003	32	10	0.74	0.70	0.906	32	5	0.56	0.57	0.876
Moorea (MO)	25	10	0.88 ± 0.04	0.005 ± 0.003	29	4	0.62	0.62	0.874	24	3	0.63	0.64	0.384
Cook Islands (CO)	27	6	0.80 ± 0.04	0.003 ± 0.002	29	5	0.57	0.57	0.918	27	3	0.48	0.65	0.065
Fiji (FI)	7	5	0.86 ± 0.14	0.004 ± 0.003	8	7	0.75	0.80	0.561	7	3	0.86	0.63	0.663
Marshall Islands (MI)	25	8	0.75 ± 0.06	0.003 ± 0.002	26	12	0.62	0.80	0.016	26	3	0.77	0.66	0.598
Palau (PL)	29	13	0.84 ± 0.06	0.004 ± 0.003	21	8	0.67	0.69	0.098	27	6	0.59	0.71	0.583
Philippines (PP)	18	8	0.80 ± 0.09	0.006 ± 0.003										
Bali (BA)	13	7	0.79 ± 0.11	0.005 ± 0.003	10	5	0.70	0.70	0.683	8	3	0.75	0.70	0.061
Christmas Island (CI)	37	13	0.68 ± 0.09	0.004 ± 0.002	32	8	0.71	0.72	0.117	36	6	0.69	0.67	0.499
Cocos-Keeling (CK)	15	7	0.86 ± 0.06	0.005 ± 0.003	17	3	0.44	0.45	1.000	16	3	0.69	0.69	0.628
<i>Acanthurus reversus</i>														
Marquesas (MQ)	30	13	0.80 ± 0.07	0.003 ± 0.002	26	6	0.73	0.68	0.610	27	3	0.56	0.58	0.183

Sample location, number of individuals (*N*), number of haplotypes (*N_h*), haplotype diversity (*h*) and nucleotide diversity (π) are listed for the cytochrome *b* (*Cytb*) data set. Number of alleles (*N_a*), observed heterozygosity (*H_O*) and expected heterozygosity (*H_E*) are listed for the *S7* ribosomal protein gene (*S7*) and the glyceraldehyde-3-phosphate dehydrogenase (*Gpd2*) nuclear introns. *P*-values are the result of exact tests for Hardy–Weinberg equilibrium using a Markov chain with 100 000 steps in ARLEQUIN. We maintained $\alpha = 0.05$ among all pairwise tests by controlling for the false discovery (Narum 2006). The corrected $\alpha = 0.015$.

generations. A sequence divergence estimate of 2% per My between lineages was used to estimate coalescence times (Bowen *et al.* 2001; Reece *et al.* 2010). Five independent runs were computed to ensure convergence, and log files were combined using the program TRACER 1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>).

Allelic states of nuclear intron sequences were estimated using the Bayesian program PHASE 2.1 (Stephens & Donnelly 2003) as implemented in the software DNASP 5.0 (Librado & Rozas 2009; see Supplemental information). Using the results from PHASE, observed heterozygosity (*H_O*) and expected heterozygosity (*H_E*) were calculated for each nuclear intron and an exact test of Hardy–Weinberg equilibrium (HWE) using 100 000 steps in a Markov chain was performed using ARLEQUIN. Linkage disequilibrium between the two nuclear loci was assessed using the likelihood ratio test with 20 000 permutations in ARLEQUIN. Genotypes for each individual at the *S7* and *Gpd2* introns were compiled and used to calculate *F_{ST}* for the multilocus data set and for pairwise comparisons between locations in ARLEQUIN. The false discovery rate among multiple comparisons was controlled as described above. Median-joining networks for alleles were constructed for each locus using NETWORK.

RADSeq data

The details of library preparation are provided in Supplemental information. Libraries were sequenced at the

UCLA Neuroscience Genomics Core facility on an Illumina HiSeq 2000 (100 bp single end reads). Sequencing runs resulted in 135 million reads passing initial quality control at the sequencing facility. Raw sequences were de-multiplexed, trimmed to a common length, and reads with Phred scores below 33 were discarded, resulting in 132 million reads with the number varying between 4.3 thousand and 4.6 million reads per individual. Nine individuals with <60 thousand reads were discarded, leaving 61 individuals across six populations (*A. olivaceus*: Hawai'i, *N* = 14; Kiritimati, *N* = 12; Marshall Islands, *N* = 7; Palau, *N* = 8; Bali, *N* = 11; *A. reversus*: Marquesas, *N* = 9). Loci were assembled using the STACKS denovo_map.pl pipeline and its component programs v. 1.09 (Catchen *et al.* 2011). Loci were generated by the merging of three or more similar 'stacks'. Stacks in which the number of reads was more than two standard deviations above the mean were assumed to be repetitive elements and removed from analysis. Loci were identified by aligning the sequence reads from each individual to the assembled contigs resulting in 844 220 stacks. Using the 'populations' component in STACKS, we further filtered the data set retaining only those loci with $\geq 8\times$ coverage ('m' command) and that aligned in $\geq 70\%$ of individuals ('r' command) in each population ('p' command). The resulting data set consisted of 3737 loci with a data matrix that was 88.5% complete. From this data set, GENEPOP 4.2 (Rousset 2008) and STRUCTURE 2.3.4 (Pritchard *et al.* 2000) input files

were produced using the 'populations' component in STACKS and implementing the 'write_single_snp' option. The resulting files were used for population level analyses or converted into other formats using the program PGDSPIDER 2.0 (Lischer & Excoffier 2012).

We tested for outlier loci across all populations using the FDI approach (Beaumont & Nichols 1996) as implemented in ARLEQUIN. This method estimates the expected distribution of Wright's inbreeding coefficient F_{ST} vs. expected heterozygosity under an island model of migration. Outlier loci are those that show either significantly higher (divergent selection) or lower (balancing selection) F_{ST} compared to simulated neutral expectations. We ran 20 000 simulations with 100 demes per group (number of groups equals the number of populations ran plus 1) and minimum and maximum expected heterozygosities of 0 and 0.8, respectively. Based on the results of this analysis, we conservatively classified each locus into one of three selection categories: neutral (between the 95% quantiles), balancing (below the 1% quantile) or divergent (above the 99% quantile) selection. Under this classification scheme, 9.7% (362) of the loci remained unclassified. We also detected outlier loci using BAYESCAN 2.1. This program uses a logistic regression model that implements a locus effect and a population effect. If the locus effect is needed to explain the observed patterns, divergent selection is considered (Foll & Gaggiotti 2008). Based on the pairwise F_{ST} values from ARLEQUIN, comparisons were made between populations of *A. olivaceus* from the Central Pacific and *A. reversus*. The results reported correspond to 100 000 simulations, with prior odds of 8 and a false discovery rate of 10%.

To test for population structure, an analysis of molecular variance (AMOVA) was performed in ARLEQUIN using 99 999 permutations. F_{ST} was calculated for the overall data set and for intraspecific pairwise comparisons among locations. Genetic partitioning was assessed using STRUCTURE with an initial burn-in of 50 000 replicates, followed by 500 000 replicates for each population (K). The analyses were run for $K = 1-6$, each replicated 10 times, without population priors. We determined the most likely number of genetic clusters (K) based on the results of STRUCTURE HARVESTER 0.6.93 (Earl & von Holdt 2012). STRUCTURE results were analysed using the program CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007), which minimizes the variance across all iterations, and the final output was visualized using the software DISTRUCT (Rosenberg 2004). Results from these preliminary runs never returned a $K > 3$; therefore, we performed 10 final runs of 500 000 replicates burn-in and 1 000 000 replicates for each of $K = 1-4$.

Finally, we conducted principle coordinate analyses (PCoA) using GENALEX 6.5 (Peakall & Smouse 2012). This

technique plots major patterns within multivariate data sets. A genetic distance matrix was calculated for each data set using the codominant (genotypic) data option. All four methods of calculation available in GENALEX (covariance-standardized, covariance-not standardized, distance-standardized and distance-not standardized) were tested and produced similar patterns. Here, we present the results of the covariance matrix without data standardization.

To resolve the function of loci under divergent selection, we searched these loci against the publicly available genomes using the NCBI Blast plug-in of GENEIOUS PRO (Biomatters). We used the programs MegaBlast and Blastn and, as expected, obtained more hits with the former. Here, we report the Blastn top hit for each fragment that had a sequence match of $\geq 50\%$ and an E -value of $\leq e-10$.

Results

Divergence of peripheral populations

We resolved 509 bp of *Cytb* in 319 individuals yielding 59 haplotypes (Fig. 2a, Table 1). Nucleotide diversity among populations was low ($\pi = 0.003-0.006$), while the corresponding haplotype diversity was high ($h = 0.51-0.88$; Table 1). Phylogenetic analysis of *Cytb* data for surgeonfishes (Acanthuridae; Table S2, Supporting information) supports the sister relationship between *A. olivaceus* and *A. reversus* (Fig. 1c and Fig. S1, Supporting information), with no diagnostic (fixed) differences between species. A median-joining network revealed a tight cluster of haplotypes, many of which are shared between putative species (Fig. 2a). Overall population structure was $\Phi_{ST} = 0.103$ ($P < 0.001$). Most of this structure was due to distinct populations at Hawai'i and the Marquesas (Table 2). Surprisingly, we found higher genetic structure distinguishing Hawai'i from conspecifics in the Central-West Pacific (pairwise $\Phi_{ST} = 0.163-0.392$) compared with population structure between *A. reversus* and central Pacific *A. olivaceus* (pairwise $\Phi_{ST} = 0.115-0.262$; Table 2). Apart from Hawai'i and the Marquesas, we detected only a single significant comparison across the remainder of the range (Christmas Island vs. Cook Islands; $\Phi_{ST} = 0.077$, $P = 0.004$).

We also resolved 171 bp of the S7 intron in 292 specimens (15 variable sites resulting in 22 alleles, Fig. 2b) and 76 bp of the Gpd2 intron in 294 specimens (five variable sites resulting in seven alleles; Fig. 2c). The number of individuals (N), number of alleles (N_a), observed heterozygosity (H_O), expected heterozygosity (H_E) and the corresponding P -value for the exact test for HWE for each locus are provided in Table 1. Despite multiple attempts, we were unable to amplify the samples from the Philip-

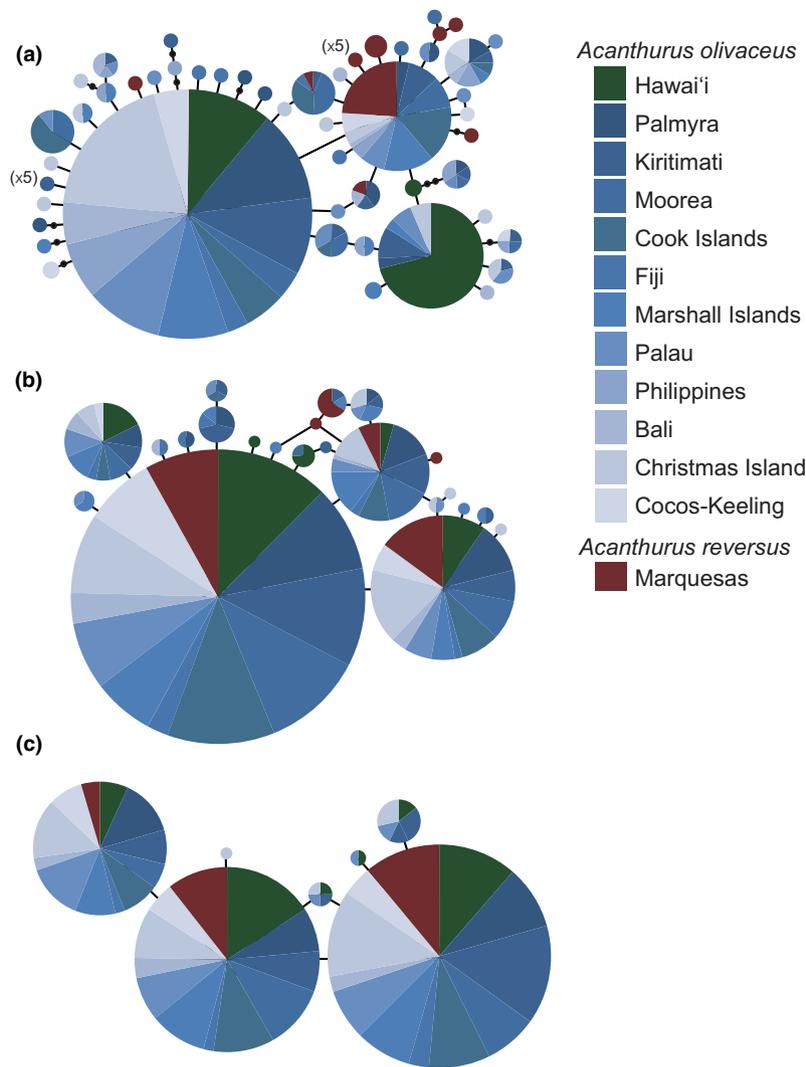


Fig. 2 Median-joining networks for (a) mtDNA cytochrome *b*, (b) the first intron of the *S7* ribosomal protein and (c) intron 2 of the Glyceraldehyde-3-phosphate dehydrogenase (*Gpd2*) gene for *Acanthurus olivaceus* and *A. reversus*. Each circle represents one mitochondrial haplotype or nuclear intron allele with the area of each circle proportional to haplotype or allele frequency; small black circles represent missing haplotypes; to simplify the figure, locations with >4 singletons branching from a single haplotype were tallied and recorded in parenthesis; colours represent collection location (see key).

Table 2 Pairwise Φ_{ST} values for *Acanthurus olivaceus* and *A. reversus* populations

	Location	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>Acanthurus olivaceus</i>	1. Hawai'i		0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2. Palmyra	<u>0.327</u>		0.216	0.023	0.030	0.828	0.468	0.538	0.671	0.724	0.449	0.499	0.000
	3. Kiritimati	<u>0.163</u>	0.009		0.262	0.024	0.342	0.674	0.974	0.613	0.438	0.286	0.393	0.000
	4. Moorea	<u>0.234</u>	0.047	0.007		0.550	0.242	0.153	0.148	0.100	0.088	0.010	0.175	0.000
	5. Cook Isl.	<u>0.376</u>	0.047	0.046	-0.008		0.225	0.081	0.031	0.032	0.077	0.004	0.136	0.000
	6. Fiji	<u>0.392</u>	-0.039	0.006	0.024	0.027		0.330	0.395	0.734	0.770	0.527	0.681	0.000
	7. Marshall Isl.	<u>0.275</u>	-0.004	-0.012	0.018	0.033	0.005		0.858	0.454	0.529	0.299	0.524	0.000
	8. Palau	<u>0.205</u>	-0.006	-0.021	0.017	0.046	0.001	-0.019		0.563	0.594	0.640	0.337	0.000
	9. Philippines	<u>0.268</u>	-0.015	-0.011	0.031	0.056	-0.033	-0.003	-0.009		0.807	0.292	0.602	0.000
	10. Bali	<u>0.309</u>	-0.017	-0.004	0.037	0.048	-0.038	-0.010	-0.013	-0.029		0.366	0.599	0.000
	11. Christmas Isl.	<u>0.271</u>	-0.003	0.004	0.059	<u>0.077</u>	-0.014	0.004	-0.010	0.006	0.000		0.113	0.000
	12. Cocos-Keeling	<u>0.294</u>	-0.006	-0.001	0.020	0.028	-0.027	-0.009	0.003	-0.011	-0.013	0.030		0.000
<i>A. reversus</i>	13. Marquesas	<u>0.397</u>	<u>0.208</u>	<u>0.161</u>	<u>0.115</u>	<u>0.158</u>	<u>0.244</u>	<u>0.166</u>	<u>0.185</u>	<u>0.192</u>	<u>0.197</u>	<u>0.262</u>	<u>0.133</u>	

Cytochrome *b* data are below diagonal, and corresponding *P*-values are above the diagonal. Significant comparisons are underlined. We maintained $\alpha = 0.05$ among all pairwise tests by controlling for the false discovery (Narum 2006). The corrected $\alpha = 0.01$.

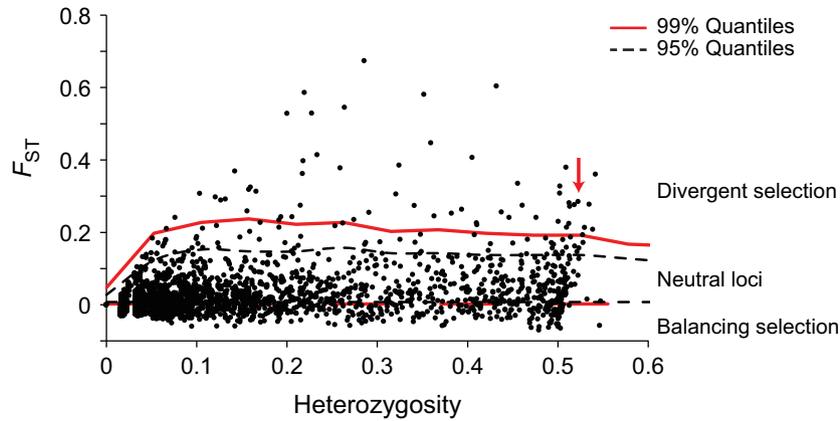


Fig. 3 Results from the FDIST analysis for the complete RADSeq data set (3737 loci). Each marker was conservatively classified into one of three categories using the F_{ST} null distribution: loci under neutral (1525 loci between the 95% quantiles), balancing (1791 loci below the 0.01% quantile, below bottom red line) or divergent (59 loci above the 99% quantile, above top red line) selection. Under this classification scheme 9.7% (362) of the loci remained unclassified. Red arrow indicates locus 65 948 (which maps to the opsin gene, Table S4, Supporting information).

Table 3 Pairwise F_{ST} values for *Acanthurus olivaceus* and *A. reversus* populations based on 3737 RADSeq loci (SNPs)

	1	2	3	4	5	6
<i>Acanthurus olivaceus</i>						
1. Hawai'i		<0.001	<0.001	<0.001	<0.001	<0.001
2. Kiritimati	<u>0.023</u>		0.065	0.486	0.011	<0.001
3. Palau	<u>0.020</u>	-0.002		0.924	0.005	<0.001
4. Marshall Isl.	<u>0.011</u>	-0.009	-0.022		0.483	<0.001
5. Bali	<u>0.027</u>	<u>0.001</u>	<u>0.001</u>	-0.009		
<i>A. reversus</i>						
6. Marquesas	<u>0.114</u>	<u>0.102</u>	<u>0.093</u>	<u>0.094</u>	<u>0.108</u>	

Corresponding P -values are above the diagonal. We maintained $\alpha = 0.05$ among all pairwise tests by controlling for the false discovery (Narum 2006). The corrected $\alpha = 0.014$. Significant comparisons are underlined.

Table 4 Pairwise F_{ST} values for RADSeq loci (SNPs) for *Acanthurus olivaceus* and *A. reversus* populations based on 1525 neutral loci (below diagonal) and 59 loci under divergent selection (above diagonal)

	1	2	3	4	5	6
<i>Acanthurus olivaceus</i>						
1. Hawai'i		<u>0.078</u>	<u>0.103</u>	<u>0.107</u>	<u>0.196</u>	<u>0.548</u>
2. Kiritimati	<u>0.029</u>		<u>0.035</u>	<u>0.046</u>	<u>0.168</u>	<u>0.552</u>
3. Palau	<u>0.032</u>	0.000		<u>0.057</u>	<u>0.114</u>	<u>0.478</u>
4. Marshall Isl.	<u>0.027</u>	0.005	-0.003		<u>0.140</u>	<u>0.526</u>
5. Bali	<u>0.033</u>	-0.003	<u>0.004</u>	<u>0.006</u>		<u>0.538</u>
<i>A. reversus</i>						
6. Marquesas	<u>0.114</u>	<u>0.098</u>	<u>0.094</u>	<u>0.101</u>	<u>0.106</u>	

We maintained $\alpha = 0.05$ among all pairwise tests by controlling for the false discovery (Narum 2006). The corrected $\alpha = 0.014$. Significant comparisons are underlined.

pinines at the nuclear introns, most likely due to template quality. Following correction for multiple comparisons, no samples significantly deviated from HWE expectations and no linkage disequilibrium was detected

between the two loci ($P > 0.05$). Similar to *Cytb*, the median-joining networks for the introns revealed tight clusters of closely related alleles that are shared between species (Fig. 2). The two most common alleles at either

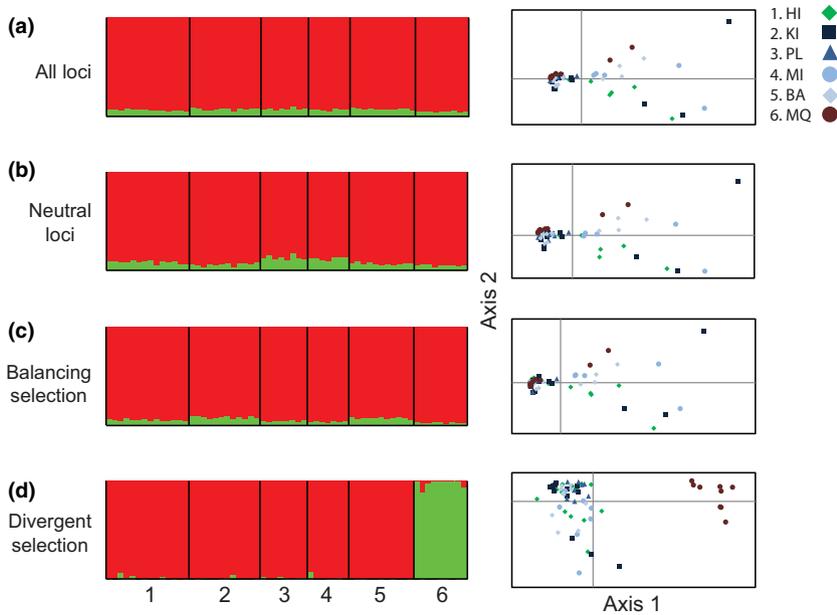


Fig. 4 Structure plots and principle coordinates analyses (PCoA). The analyses were run using (a) all RADSeq loci (3737), (b) neutral loci (1525), (c) loci classified under balancing selection (1791) and (d) loci classified under divergent selection (59). Abbreviations: Hawai'i (HI), Kiritimati (KI), Palau (PL), Marshall Islands (MI), Bali (BA) and Marquesas Islands (MQ).

locus were observed in both species and at every location, except Cocos-Keeling (*S7* locus). We detected no population structure in the combined (multi-locus) nuclear intron data set (overall $F_{ST} = 0.002$, $P = 0.185$), and none of the pairwise comparisons were significant after correction for multiple comparisons (data not shown). When the introns were evaluated separately, overall population structure in the *S7* data set was low but significant (overall $F_{ST} = 0.009$, $P = 0.038$) with only a single significant pairwise comparison (Hawai'i vs. Marquesas, $F_{ST} = 0.048$, $P = 0.005$; Table S3, Supporting information). No structure was detected with *Gpd2* intron (overall $F_{ST} = 0.009$, $P = 0.038$; Table S3, Supporting information).

Coalescent times

To estimate divergence time, we conducted coalescent analysis on the complete mtDNA data set, which resulted in an estimated time to most recent common ancestor (TMRCA) for *A. olivaceus* and *A. reversus* of 0.83 Ma (95% HPD = 0.45–1.36 Ma). When analysed separately, TMRCA for *A. olivaceus* was 0.68 Ma (95% HPD = 0.41–0.99 Ma) and 0.49 Ma for *A. reversus* (95% HPD = 0.28–0.77 Ma). When only the Hawaiian population of *A. olivaceus* was considered, TMRCA = 0.55 Ma (95% HPD = 0.34–0.79 Ma). Central-West Pacific populations of *A. olivaceus* were similar (TMRCA = 0.58 Ma, 95% HPD = 0.38–0.82 Ma). While caution is warranted because these dates were calculated based on a single mtDNA marker, they imply that *A. reversus* began diverging from the Central-West Pacific population (0.49 Ma) at about the same time as the Hawaiian population of *A. olivaceus* (0.55 Ma).

Genome scans

We amplified 3737 RADSeq loci in 61 individuals across six sample locations (Fig. 1b) producing a robust data matrix that was 88.5% complete. Using the 'write_single_snp' command in STACKS (as described above), we analysed the first single nucleotide polymorphism (SNP) in each locus. Employing the FDIST approach across all populations (ARLEQUIN), we detected 1525 neutral loci, 1791 loci under balancing selection, and 59 loci under divergent selection (pairwise $F_{ST} = 0.035$ –0.552; Table 4, Fig. 3). Under this conservative classification scheme, 9.7% (362) of the loci remained unclassified. We detected 16 outlier loci using BAYESCAN, which were a subset of the ones detected using the FDIST approach with the exception of a single locus (Table S4, Supporting information). When the Marquesas population was removed from the analysis, the number of loci under divergent selection detected using the FDIST approach dropped by 43% (to 34 loci) with lower F_{ST} values (pairwise $F_{ST} = 0.14$ –0.39, data not shown). For all subsequent analyses, we used the 59 outlier loci detected using the FDIST approach. Similar to the mtDNA data set, we found significant population structure across the RADSeq loci (overall $F_{ST} = 0.039$, $P < 0.001$). However, this larger data set revealed differences between the putative species that are concordant with expectations based on colour morphology. While the Hawaiian population of *A. olivaceus* is genetically distinct from conspecific populations in the Central-West Pacific (pairwise $F_{ST} = 0.011$ –0.027, Table 3), structure between *A. reversus* and Central-West Pacific populations of *A. olivaceus* was higher (pairwise $F_{ST} = 0.093$ –0.108). In

contrast to the mtDNA data set we detected low but significant genetic structure in the population of *A. olivaceus* at Bali (Table 3). When the loci under balancing selection were considered, we found no population structure (overall $F_{ST} = -0.031$, $P = 1.000$). In contrast, when the neutral loci were analysed, we found significant population structure ($F_{ST} = 0.043$, $P < 0.001$; Table 4). Pairwise F_{ST} values were highest for the Marquesan population ($F_{ST} = 0.094$ – 0.106) and lower for Hawai'i ($F_{ST} = 0.027$ – 0.033). Bali demonstrated low but significant structure when compared to both Palau ($F_{ST} = 0.004$, $P < 0.001$) and the Marshall Islands ($F_{ST} = 0.006$, $P < 0.001$). For loci under divergent selection, overall F_{ST} was high and significant ($F_{ST} = 0.313$, $P < 0.001$) and all pairwise F_{ST} values were significant. Population structure was highest for the Marquesan population (pairwise $F_{ST} = 0.478$ – 0.552), while values from Hawai'i (pairwise $F_{ST} = 0.078$ – 0.196) and within the Central-West Pacific were much lower (pairwise $F_{ST} = 0.035$ – 0.168 ; Table 4).

Results from the STRUCTURE analyses indicate $K = 2$ (3737 loci) with a lack of clustering among populations that is supported by the PCoA (Fig. 4a). Analyses of the neutral loci (Fig. 4b, $K = 2$) and the loci under balancing selection (Fig. 4c, $K = 2$ and 3 had similar probabilities and lack of clustering; $K = 2$ is shown) resulted in similar patterns with no partitioning among populations. In contrast, when the 59 outlier loci were analysed, STRUCTURE resulted in $K = 2$ with *A. reversus* clearly distinguished from *A. olivaceus*, but with no corresponding distinction of the Hawaiian population. The distinction of the Marquesan population was supported by PCoA only when the outlier loci were considered (Fig. 4d). The results of our NCBI Blast searches mapped one of the outlier loci to the opsin Rh2 gene (retinal pigment Rh 2A alpha; locus 65 948; Table S4, Supporting information). There were two alleles detected at this locus that are in nearly equal frequency across the data set (average gene diversity = 0.49 ± 0.47). However, allele frequencies at the Marquesas were nearly fixed for one variant (90%).

Discussion

While we have a theoretical understanding of the roles that natural selection and geographic isolation play in shaping biodiversity, empirical studies of these processes in the sea are rare. The Orangeband Surgeonfish species complex forms a monophyletic lineage (Fig. 1c) and provides an unusual opportunity to address the role of selection in speciation with one geographically isolated population in the Hawaiian Islands and a sister species (*A. reversus*) of similar age (0.55 and 0.49 Ma), endemic to the Marquesas Islands. Using a targeted

locus approach (mtDNA and nuclear introns), we found no diagnostic differences between these two lineages (Fig. 2). However, we did observe a strong shift in mtDNA haplotype frequencies (Table 2) with no accompanying signal of divergence in nuclear introns (Table S3, Supporting information). While little population structure was detected among *A. olivaceus* populations across the Central-West Pacific, the remote Hawaiian Archipelago was distinct and demonstrated population genetic divergence in mtDNA (shifts in haplotype frequencies) similar to *A. reversus* in the Marquesas Islands (Fig. 2 and Table 2). In contrast, our genome-wide survey revealed low levels of population structure at Bali not detected in the Sanger data sets (Tables 3 and 4). Most interesting was the detection of signals of divergent selection that distinguish the Marquesas, a signal not observed in the Hawaiian population of *A. olivaceus*, but with dozens of loci under divergent selection distinguishing *A. reversus*. This pattern is in concordance with the expectations of geographical isolation of Hawai'i, versus divergent selection acting upon lineages in the Marquesas.

Our data set does not negate the role of selection in geographically isolated populations such as those in Hawai'i nor does it dismiss the role of drift in populations under strong selection. On the contrary, our data set indicates that drift at neutral loci is important in both populations (Table 4), but it also highlights the significance of selection in disparate environments such as those observed in the Marquesas, adding to the accumulating evidence that selection in contrasting environments can be an important driver of speciation in the sea.

Isolation in Hawai'i and selection at the Marquesas

As expected from traditional views of allopatric speciation, much of the high rates of endemism in the oceans are observed in geographically isolated locations, such as oceanic islands (Briggs & Bowen 2012). The Hawaiian Archipelago, which has the highest level of endemism in Oceania (25% of shore fishes; Randall 2007), is one of the most remote island groups on Earth. Hawai'i is isolated to the east by a 4000 km stretch of open ocean. While Johnston Atoll lies just 850 km to the southwest, this island is small (2.7 km²) and provides little habitat. The small islands and atolls of the Line Islands lie about 1600 km to the south of Hawai'i but their northernmost atolls, Kingman Reef and Palmyra, are swept by equatorial currents that flow in a predominantly east–west direction. As a result, Hawai'i is exceptionally isolated and colonization of this archipelago likely results from infrequent dispersal, yielding high endemism and genetically distinct

populations (Craig *et al.* 2010; Bird *et al.* 2011; Gaither *et al.* 2011).

Similar to Hawai'i, the Marquesas are also isolated to the east. However, just 475 km southwest of the Marquesas lies a network of islands that extends to Australia, with distances no greater than 800 km between shallow habitats (Schultz *et al.* 2008). Prevailing currents in the region flow in a westerly direction, but genetic studies indicate that dispersal in the opposing direction is frequent enough to preclude the development of population structure among some reef fishes (Gaither *et al.* 2010). Thus, the high rate of endemism at these islands is difficult to explain by geographic isolation alone. Instead, what distinguish the Marquesas are unique environmental conditions. Unlike most tropical Pacific islands, the Marquesas lack fringing reefs and lagoons. These islands are exposed to unusually variable sea temperatures for an equatorial location with temperatures ranging from 26 to 30 °C in the open ocean, while there are reports of temperatures as low 15–18 °C due to upwelling (Randall & Earle 2000). Additionally, there are frequent phytoplankton blooms and restricted coral development (Randall & Earle 2000; Martinez & Maamaatuaiahutapu 2004). The lack of lagoon habitat and reduced coral cover, combined with variable sea water temperatures and high nutrient input, is thought to drive the biogeographic distinction of the Marquesas, which harbours unusually low species richness and high endemism (12% of shore fishes; Randall 2005). Furthermore, several phylogeographic surveys of widely distributed species have shown significant population differentiation between the Marquesas and the Central Pacific (Planes & Fauvelot 2002; Gaither *et al.* 2010; Ludt *et al.* 2012; Szabó *et al.* 2014). Colonists that arrive here are likely to be under strong selection in these unique ecological conditions, which may lead to differentiation of lineages and, ultimately, to speciation.

Colour as a taxonomic trait

The delineation of species is fundamental to many aspects of evolutionary biology, systematics, biogeography and conservation. Until recently, most species were distinguished based on morphology; however, molecular technologies arising in the past few decades have revealed the shortcomings of using morphology as the sole criteria for taxonomy. An exceptional example is the bonefish *Albula vulpes*. Based on morphology, this species was considered a single globally distributed species until analysis of molecular data revealed deep genetic partitions in Hawaiian populations (Shaklee & Tamaru 1981). More recent analyses indicate the existence of 12 or more cryptic lineages within *A. vulpes*

with sister lineages separated by at least 3.5 Ma (Bowen *et al.* 2007; Hidaka *et al.* 2008).

Colour is often used to delineate closely related taxa especially among the vibrantly coloured species that inhabit tropical coral reefs. However, discordance between genetic divergence and coloration is well documented in butterflyfishes (McMillan & Palumbi 1995) and hamlets (Ramon *et al.* 2003; Garcia-Machado *et al.* 2004). The brightly coloured pygmy angelfishes (Genus *Centropyge*) offer striking examples. The Flame Angelfish (*Centropyge loricula*) maintains distinct geographic colour morphs across its Central Pacific range without corresponding genetic partitioning at neutral loci (Schultz *et al.* 2007). *Centropyge* in the Atlantic have largely allopatric distributions and heritable colour patterns, yet they are phylogenetically indistinguishable (Bowen *et al.* 2006). Three species have been described in the *Centropyge flavissima* complex based on colour; however, distinct mtDNA lineages cluster by geographic location (Pacific Ocean, Indian Ocean and Moorea) and not by species (DiBattista *et al.* 2012). In the Blackstreak Cleaner Wrasse *Labroides dimidiatus*, mtDNA lineages group by Pacific archipelago and not colour morph (Sims *et al.* 2013).

Acanthurus reversus was described based on diagnostic colour differences: the orange band in *A. reversus* is much smaller than in *A. olivaceus*, and the tail has a 'reverse' coloration, with a white margin and a black submarginal band, whereas in *A. olivaceus*, the margin is black and the submarginal band is white (hence the name, *Acanthurus reversus*; Fig. 1a). Analyses of mtDNA and nuclear introns indicate that this colour change is not accompanied by reciprocal monophyly (Figs 1c and 2). Only with the analysis of the much larger RADSeq data set is *A. reversus* distinguished from its sister taxa with a strong signal of selection (Fig. 4). Coalescent analysis indicates that evolution of colour in *A. reversus* was rapid (TMRCA = 0.83 Ma) relative to the species complexes of *Centropyge* described above (TMRCA = 2.7–12 Ma, Gaither *et al.* 2014). Taken together, these studies confirm that colour evolves at different rates across groups, does not necessarily correspond with the evolutionary partitions detected in phylogenetic studies and should be used as a taxonomic trait with caution.

Selection on rhodopsin genes

Our RADSeq analysis revealed 59 loci (of 3737) under divergent selection, with one of them (locus 65 948; Table S4, Supporting information) mapping to the opsin Rh2 gene. Opsins are a group of light-sensitive pigments involved in mediating the conversion of light into an electrochemical signal, the first step in the visual transduction cascade. The Rh2 gene is an essential com-

ponent of the cone receptor system that is sensitive to middle wavelengths of the visible spectrum (Owens *et al.* 2009).

Coral reefs are generally characterized by clear oligotrophic waters (Birkeland 1988); however, coral reefs in the Marquesas are atypical. Episodic increases in nutrients around the islands, attributed to a combination of ocean currents, El Niño events, and terrestrial run-off, yield algal blooms and lower visibility (Randall & Earle 2000; Martinez & Maamaatuaiahutapu 2004). Selection in opsin genes has been detected in a number of species, including African cichlids, where fixation of different alleles has been shown to correspond to adaptation to different light regimes and coincides with divergence in male breeding coloration (Spady *et al.* 2005; Sugawara *et al.* 2005; Terai *et al.* 2006). Further, positive selection on Rh1 and Rh2 opsin genes may have been important in the early evolution of the group (Dann *et al.* 2004). In marine gobies (family Gobiidae), amino acid substitutions at spectral turning sites of the Rh1 gene correspond with divergent light environments related to depth, turbidity and light spectrum (Larmuseau *et al.* 2010). In the Cormoran coelacanths (*Latimeria chalumnae*), amino acid replacements at opsin genes have been tied to adaptation to the low light conditions at depth (Yokoyama *et al.* 1999). Taking these examples into account, the low light conditions that characterize the Marquesas could be driving the selection in visual pigments detected here in *A. reversus*. Future studies targeting opsin genes across taxonomic groups will help inform the role of these pigments in adaptation of fishes at the Marquesas.

The significant shifts in haplotype frequency detected in *A. reversus*, combined with a dramatic divergence in colour and signals of selection that correspond with known biogeographic partitions, indicate *A. reversus* is on an independent evolutionary trajectory. While the lack of monophyly may be unsatisfactory to some, we prefer to avoid unnecessary nomenclature juggling and therefore suggest no change in taxonomy.

The utility of RADSeq in phylogeographic studies

Over the past few decades, the number of mtDNA studies has grown rapidly and the inferences drawn from these data sets have greatly improved our understanding of phylogeography, evolution and ecology. From these data, we can calibrate molecular clocks across a diversity of marine taxa (Bowen *et al.* 2001; Lessios 2008; Reece *et al.* 2010). However, gene trees generated from the maternally inherited mtDNA genome, while biologically informative, may not represent the species evolutionary history. The alternative of producing nuclear intron sequence data can be time-con-

suming and expensive, which limits the number of loci (and the proportion of the genome) analysed. Moreover, sequencing approaches that target neutral loci are unlikely to reveal signals of selection.

With new and user-friendly laboratory protocols (Peterson *et al.* 2012; Toonen *et al.* 2013) and bioinformatics pipelines (Catchen *et al.* 2011; Puritz *et al.* 2014), as well as the falling price of sequencing, we can now resolve thousands of loci across the genomes of nonmodel organisms quickly and cost effectively. The utility of this approach, as well as others that target specific genomic regions (i.e. target enrichment or sequence capture), is changing the way we conduct phylogeographic studies (McCormack *et al.* 2013; Rocha *et al.* 2013; Andrews & Luikart 2014; Bowen *et al.* 2014). While the nuclear introns assayed here showed a near complete lack of genetic structure, the RADSeq data revealed low levels of population structure at Bali (Tables 3 and 4). The latter may be due to the position of Bali at the western end of the species range. Perhaps the most interesting aspect is, however, the ability to detect divergent selection at the Marquesas, a finding only made possible with the amplification of 1000s of loci randomly distributed across the genome.

Conclusion

Phylogenetic and phylogeographic analyses of neutral loci may lack the resolution to detect recent isolation or rapid divergence. The distinct ecological conditions at the Marquesas, and the finding of genetic divergence across numerous taxa, indicate that both selection and genetic drift are driving divergence at these islands. The detection of 59 outlier loci, combined with divergence in coloration and a signature of selection in visual pigment genes, indicates that *A. reversus* is on an independent evolutionary trajectory, driven by divergent selection in the unique environment at the Marquesas. The synchronous isolation in Hawai'i and the Marquesas demonstrates how adaptation can accelerate the speciation process. In the half million years since these populations were founded, the geographic isolation of Hawai'i has produced a divergent population, whereas the unique ecology of the Marquesas has produced what taxonomists recognize as a distinct species. This is the first case study of its kind for coral reef fishes, and we believe that further studies will confirm that speciation driven by natural selection is an important source of biodiversity in the sea.

Acknowledgements

This study was supported by the National Science Foundation (NSF) under grants No. OIA0554657 and OCE-0929031 (BWB).

Additional support for this work came from NOAA National Marine Sanctuaries Program MOA grant No. 2005-008/66882 (R.J. Toonen), Hawai'i Sea Grant under grant No. NA05O AR4171048 (BWB) and the California Academy of Sciences (LAR). Support for MAB provided by SENACYT. Fieldwork at Cocos-Keeling and Christmas Island was supported by National Geographic Grant 8208-07 (M.T. Craig). Thanks to Rob Toonen, Howard Choat, Pat Colin, Lori Colin, Joshua Copus, Matthew Craig, Toby Daly-Engel, Joseph DiBattista, Matthew Dunlap, John Earle, Jeff Eble, Brian Greene, Jean-Paul Hobbs, Matthew Iacchi, Stephen Karl, Randall Kosaki, Carl Meyer, Yannis Papastamatiou, David Pence, Richard Pyle, John Randall, Joshua Reece, D. Ross Robertson, Ben Rome, Jennifer Schultz, Charles Sheppard, Craig Skepper, Zoltan Szabo, Kim Tenggardjaja, Tukabu Teroroko, Robert Thorne, Daniel Wagner, Robert Whitton, Jill Zamzow, the Coral Reef Research Foundation, and the Papanumokuakea Marine National Monument for specimens, advice and logistical support, and Kydd Pollock and Zachary Caldwell at the research station on Palmyra and especially Erik Conklin at TNC and Amanda Meyer Pollock at the US Fish and Wildlife Service Palmyra for making collections there possible. We thank Anna Sellas at the California Academy of Sciences' Center for Comparative Genomics for RAD-Seq protocol training. We also thank Claudia Rocha, Durrell Kapan, and John Randall for logistical support and lively discussions. Thanks to John Randall and Keoki Stender for photographs. We thank the staff of the University of Hawai'i Advanced Studies of Genomics, Proteomics and Bioinformatics sequencing facility, the Hawai'i Institute of Marine Biology, and the California Academy of Sciences for their support throughout this project. This is contribution #1615 from the Hawai'i Institute of Marine Biology, #9270 from the UH School of Ocean and Earth Science and Technology, and UNIHI-SEAGRANT-JC-11-04 from University of Hawai'i Sea Grant.

References

- Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. *Molecular Ecology*, **23**, 1661–1667.
- Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.
- Bird CE, Holland BS, Bowen BW, Toonen RJ (2011) Diversification of sympatric broadcast-spawning limpets (*Cellana* spp.) within the Hawaiian archipelago. *Molecular Ecology*, **20**, 2128–2141.
- Birkeland C (1988) Geographic comparisons of coral-reef community processes. In: *Proceedings of the Sixth International Coral Reef Symposium* (eds Choat JH, Barnes D, Borowitzka MA, Coll JC, Davies PJ, Flood P, Hatcher BG & Hopley D), pp. 211–220. Townsville, Australia.
- Blais J, Plenderleith M, Rico C *et al.* (2009) Assortative mating among Lake Malawi cichlid fish populations is not simply predictable from male nuptial colour. *BMC Evolutionary Biology*, **9**, 53.
- 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.
- Bowen BW, Muss A, Rocha LA, Grant WS (2006) Shallow mtDNA coalescence in Atlantic pygmy angelfish (genus *Centropyge*) indicates a recent invasion from the Indian Ocean. *Journal of Heredity*, **97**, 1–12.
- Bowen BW, Karl SA, Pfeiler E (2007) Resolving evolutionary lineages and taxonomy of bonefishes (*Albula* spp.). In: *Biology and Management of the World Tarpon and Bonefish Fisheries* (ed. Ault JS), pp. 147–154. CRC Press, Boca Raton, Florida.
- Bowen BW, Rocha LA, Toonen RJ *et al.* (2013) The origins of tropical marine biodiversity. *Trends in Ecology and Evolution*, **28**, 359–366.
- Bowen BW, Shanker K, Yasuda N *et al.* (2014) Phylogeography unplugged: comparative surveys in the genomic era. *Bulletin of Marine Science*, **90**, 13–46.
- Briggs JC, Bowen BW (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. *Journal of Biogeography*, **39**, 12–30.
- Briggs JC, Bowen BW (2013) Evolutionary patterns: marine shelf habitat. *Journal of Biogeography*, **40**, 1023–1035.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.
- Choat JH, Axe LM (1996) Growth and longevity in acanthurid fishes: an analysis of otolith increments. *Marine Ecology Progress Series*, **134**, 15–26.
- Choat JH, Klanten OS, Van Herwerden L, Robertson DR, Clements KD (2012) Patterns and processes in the evolutionary history of parrotfishes (Family Labridae). *Biological Journal of the Linnean Society*, **107**, 529–557.
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts, pp. 545.
- Craig MT, Hastings PA, Pondella DJ, Robertson DR, Rosales-Casian JA (2006) Phylogeography of the flag cabrilla *Epinephelus labriformis* (Serranidae): implications for the biogeography of the Tropical Eastern Pacific and the early stages of speciation in a marine shore fish. *Journal of Biogeography*, **33**, 969–979.
- Craig MT, Eble J, Bowen BW (2010) Origins, ages, and population histories: comparative phylogeography of endemic Hawaiian butterflyfishes (genus *Chaetodon*). *Journal of Biogeography*, **37**, 2125–2136.
- Dann SG, Allison WT, Levin DB, Taylor JS, Hawryshyn CW (2004) Salmonid opsin sequences undergo positive selection and indicate an alternate evolutionary relationship in *Oncorhynchus*. *Journal of Molecular Evolution*, **58**, 400–412.
- Darwin CR (1859) *On The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 6th edn. John Murray, London, UK.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.
- DiBattista JD, Waldrop E, Bowen BW *et al.* (2012) Twisted sister species of pygmy angelfishes: discordance between taxonomy, coloration, and phylogenetics. *Coral Reefs*, **31**, 839–851.
- Doherty PJ, Planes S, Mather P (1995) Gene flow and larval duration in seven species of fish from the Great Barrier Reef. *Ecology*, **76**, 2373–2391.

- Drew J, Barber PH (2009) Sequential cladogenesis of the reef fish *Pomacentrus moluccensis* (Pomacentridae) supports the peripheral origin of marine biodiversity in the Indo-Australian archipelago. *Molecular Phylogenetics and Evolution*, **53**, 335–339.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
- Endler JA, Houde AE (1995) Geographic variation in female preferences for male traits in *Poecilia reticulata*. *Evolution*, **49**, 456–468.
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Fisher R, Leis JM, Clark DL, Wilson SK (2005) Critical swimming speeds of late-stage coral reef fish larvae: variation within species, among species and between locations. *Marine Biology*, **147**, 1201–1212.
- 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**, 977–993.
- Gaither MR, Rocha LA (2013) Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. *Journal of Biogeography*, **40**, 1638–1648.
- Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW (2010) Genetic evaluation of marine biogeographic barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). *Journal of Biogeography*, **37**, 133–147.
- Gaither MR, Jones SA, Kelley C, Newman SJ, Sorenson L, Bowen BW (2011) High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian archipelago. *PLoS ONE*, **6**, e28913.
- Gaither M, Schultz JK, Bellwood DR *et al.* (2014) Evolution of pygmy angelfishes: recent divergences, introgression, and the usefulness of color in taxonomy. *Molecular Phylogenetics and Evolution*, **74**, 38–47.
- Garcia-Machado E, Chevalier Monteagudo PP, Solignac M (2004) Lack of mtDNA differentiation among hamlets (*Hypoplectrus*, Serranidae). *Marine Biology*, **144**, 147–152.
- Helfman GS, Collett BB, Facey DE, Bowen BW (2009) *The Diversity of Fishes: Biology, Evolution, and Ecology*, 2nd edn. Wiley-Blackwell, Oxford, UK.
- Hidaka K, Iwatsuki Y, Randall JE (2008) A review of the Indo-Pacific bonefishes of the *Albula argentea* complex, with a description of a new species. *Ichthyological Research*, **55**, 53–64.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.
- 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**, 1801–1806.
- Johannesson K (2009) Inverting the null-hypothesis of speciation: a marine snail perspective. *Evolutionary Ecology*, **23**, 5–16.
- Larmuseau MHD, Huysse T, Vancampenhout K, Van Houdt JKJ, Volckaert FAM (2010) High molecular diversity in the rhodopsin gene in closely related goby fishes: a role for visual pigments in adaptive speciation? *Molecular Phylogenetics and Evolution*, **55**, 689–698.
- Leis JM (1991) The pelagic phase of coral reef fishes: larval biology of coral reef fishes. In: *The Ecology of Fishes on Coral Reefs* (ed. Sale PF), pp. 183–230. Academic Press, San Diego.
- Leis JM, McCormick MI (2002) The biology, behavior, and ecology of the pelagic, larval stage of coral reef fishes. In: *Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem* (ed. Sale PF), pp. 171–199. Academic Press, San Diego.
- Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G (2010) Allopatric divergence and speciation in coral reef fish: the three-spot *Dascyllus*. *Dascyllus trimaculatus*, species complex. *Evolution*, **64**, 1218–1230.
- Lessios HA (2008) The great American schism: divergence of marine organisms after the rise of the Central American Isthmus. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 63–91.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.
- Ludt WB, Bernal MA, Bowen BW, Rocha LA (2012) Living in the past: phylogeography and population histories of Indo-Pacific wrasses (genus *Halichoeres*) in shallow lagoons versus outer reef slopes. *PLoS ONE*, **7**, e38042.
- Martinez E, Maamaatuaiahutapu K (2004) Island mass effect in the Marquesas Islands: time variation. *Geophysical Research Letters*, **31**, L18307.
- McCartney MA, Acevedo J, Heredia C *et al.* (2003) Genetic mosaic in a marine species flock. *Molecular Ecology*, **12**, 2963–2973.
- McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.
- McMillan WO, Palumbi SR (1995) Concordant evolutionary patterns among Indo-West Pacific butterflyfishes. *Proceedings of the Royal Society B: Biological Sciences*, **260**, 229–236.
- McMillan WO, Weigt LA, Palumbi SR (1999) Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). *Evolution*, **53**, 247–260.
- Messmer V, van Herwerden L, Munday PL, Jones GP (2005) Phylogeography of colour polymorphism in the coral reef fish *Pseudochromis fuscus*, from Papua New Guinea and the Great Barrier Reef. *Coral Reefs*, **24**, 392–402.
- Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9724–9729.
- Munday PL, Van Herwerden L, Dudgeon CL (2004) Evidence for sympatric speciation by host shift in the sea. *Current Biology*, **14**, 1498–1504.
- Narum SR (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics*, **7**, 783–787.

- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York City, New York.
- Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford, UK.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.
- Owens GL, Windsor DJ, Mui J, Taylor JS (2009) A fish eye out of water: ten visual opsins in the four-eyed fish, *Anableps anableps*. *PLoS ONE*, **4**, e5970.
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **28**, 2537–2539.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.
- Planes S, Fauvelot C (2002) Isolation by distance and vicariance drive genetic structure of a coral reef fish in the Pacific Ocean. *Evolution*, **56**, 378–399.
- Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.
- Prada C, Hellberg ME (2013) Long prereproductive selection and divergence by depth in a Caribbean candelabrum coral. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 3961–3966.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Puebla O, Bermingham E, Guichard F, Whiteman E (2007) Colour pattern as a single trait driving speciation in *Hypoplectrus* coral reef fishes? *Proceedings of the Royal Society B: Biological Sciences*, **274**, 1265–1271.
- Puebla O, Bermingham E, Guichard F (2011) Pairing dynamics and the origin of species. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1085–1092.
- Puebla O, Bermingham E, McMillan WO (2014) Genomic atolls of differentiation in coral reef fishes (*Hypoplectrus* spp, Serranidae). *Molecular Ecology*, doi: 10.1111/mec.12926.
- Puritz J, Hollenbeck CM, Gold JR (2014) *dDocent*: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, **2**, e314v1.
- Ramon ML, Lobel PS, Sorenson MD (2003) Lack of mitochondrial genetic structure in hamlets (*Hypoplectrus* spp.): recent speciation or ongoing hybridization? *Molecular Ecology*, **12**, 2975–2980.
- Randall JE (2005) *Reef and Shore Fishes of the South Pacific*. Sea Grant College Program, University of Hawai'i Press, Honolulu, Hawaii.
- Randall JE (2007) *Reef and Shore Fishes of the Hawaiian Islands*. Sea Grant College Program, University of Hawai'i Press, Honolulu, Hawaii.
- Randall JE, Earle JL (1999) *Acanthurus reversus*, a new species of surgeonfish (Perciformes: Acanthuridae) from the Marquesas Islands. *Proceedings of the California Academy of Sciences*, **51**, 473–481.
- Randall JE, Earle JL (2000) Annotated checklist of the shore fishes of the Marquesas Islands. *Bishop Museum Occasional Papers*, **66**, 1–42.
- Reece JS, Bowen BW, Smith DG, Larson AF (2010) Molecular phylogenetics of moray eels (Muraenidae) demonstrates multiple origins of a shell-crushing jaw (*Gymnomuraena*, *Echidna*) and multiple colonizations of the Atlantic Ocean. *Molecular Phylogenetics and Evolution*, **57**, 829–835.
- Renaut S, Maillet N, Normandeau E et al. (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 354–363.
- Robertson DR (1983) On the spawning behavior and spawning cycles of eight surgeonfishes (Acanthuridae) from the Indo-Pacific. *Environmental Biology of Fishes*, **9**, 193–223.
- Rocha LA (2004) Mitochondrial DNA and color pattern variation in three western Atlantic *Halichoeres* (Labridae), with the revalidation of two species. *Copeia*, **2004**, 770–782.
- Rocha LA, Robertson DR, Roman J, Bowen BW (2005) Ecological speciation in tropical reef fishes. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 573–579.
- Rocha LA, Bernal MA, Gaither MR, Alfaro M (2013) Massively parallel DNA sequencing: the new frontier in biogeography. *Frontiers in Biogeography*, **5**, fb_16189.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.
- Salzburger W, Niedersttter H, Brandstatter A et al. (2006) Colour-assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 257–266.
- Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372–380.
- Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.
- Schultz JK, Pyle RL, DeMartini E, Bowen BW (2007) Genetic connectivity among color morphs and Pacific archipelagos for the Flame Angelfish, *Centropyge loriculus*. *Marine Biology*, **151**, 167–175.
- Schultz JK, Feldheim KA, Gruber SH, Ashley MV, McGovern TM, Bowen BW (2008) Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*). *Molecular Ecology*, **17**, 5336–5348.
- Selz OM, Pierotti MER, Mann ME, Schmid C, Seehausen O (2014) Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behavioral Ecology*, **25**, 612–626.
- Shaklee JB, Tamaru CS (1981) Biochemical and morphological evolution of Hawaiian bonefishes (*Albula*). *Systematic Zoology*, **30**, 125–146.
- Sims CA, Riginos C, Blomberg SP, Huelsenken T, Drew J, Grutter AS (2013) Cleaning up the biogeography of *Labroides dimidiatus* using phylogenetics and morphometrics. *Coral Reefs*, **33**, 223–233.
- Sorenson L, Santini F, Carnevale G, Alfaro M (2013) A multi-locus timetree of surgeonfishes (Acanthuridae, Percomor-

- pha), with revised family taxonomy. *Molecular Phylogenetics and Evolution*, **68**, 150–160
- Spady TC, Seehausen O, Loew ER, Jordan RC, Kocher TD, Carleton KL (2005) Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Molecular Biology and Evolution*, **22**, 1412–1422
- Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169
- Sugawara T, Terai Y, Imai H *et al.* (2005) Parallelism of amino acid changes at the Rh1 affecting spectral sensitivity among deep-water cichlids from Lakes Tanganyika and Malawi. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 5448–5453
- Supple M, Papa R, Counterman B, McMillan WO (2014) The genomics of an adaptive radiation: insights across the *Heliconius* speciation continuum. *Advances in Experimental Medicine and Biology*, **781**, 249–271
- Szabó Z, Snelgrove B, Craig MT, Rocha LA, Bowen BW (2014) Phylogeography of the Manybar Goatfish, *Parupeneus multifasciatus* reveals moderate structure between the Central and North Pacific and a cryptic endemic species in the Marquesas. *Bulletin of Marine Science*, **90**, 493–512
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2793
- Terai Y, Seehausen O, Sasaki T *et al.* (2006) Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biology*, **4**, e433
- Toonen RJ, Puritz JB, Forsman ZH *et al.* (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, **1**, e203
- Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9939–9946
- Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London*, **367**, 451–460
- Wu C (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865
- Yokoyama S, Zhang H, Radlwimmer FB, Blow NS (1999) Adaptive evolution of color vision of the Comoran coelacanth (*Latimeria chalumnae*). *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 6279–6284
-
- M.R.G., B.W.B. and L.A.R. designed the project and supervised field operations. M.R.G., M.A.B., B.W.B. and L.A.R. wrote the initial manuscript. M.R.G., M.A.B., R.R.C. and S.A.J. produced Sanger data and prepared RADSeq libraries. W.B.S. performed quality control of RADSeq data and provided bioinformatics support. M.R.G., M.A.B. and R.R.C. analysed data. All authors contributed to editing and revising the manuscript.
-

Data accessibility

Sanger sequence alignments and RADSeq raw data were deposited in Dryad, entry doi: 10.5061/dryad.581f3.

Supporting information

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

Fig. S1 Phylogenetic tree for cytochrome *b* sequences of Acanthurid species for which sequences (> 600bp) were publicly available (Table S2, Supporting information).

Table S1 Morphometrics for species of *Acanthurus* from Randall (2005).

Table S2 List of species and GenBank accession numbers for cytochrome *b* sequences used in phylogenetic analyses (Fig. S1, Supporting information).

Table S3 Pairwise F_{ST} values for eleven populations of *Acanthurus olivaceus* and *A. reversus* from the Marquesas.

Table S4 Blastn top hit for each of 59 outlier loci with a sequence match of higher than $\geq 50\%$ and an *E*-value of $\leq e-10$ or lower.

Appendix S1 Materials and methods.