MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 2902–2913 doi: 10.1111/mec.12782

Seascape continuity plays an important role in determining patterns of spatial genetic structure in a coral reef fish

C. C. D'ALOIA,* S. M. BOGDANOWICZ,† R. G. HARRISON† and P. M. BUSTON*

Department of Biology and Marine Program, Boston University, 5 Cummington Mall, Boston, MA 02215, USA, †Department of Ecology and Evolutionary Biology, Cornell University, Corson Hall, Ithaca, NY 14853, USA

Abstract

Detecting patterns of spatial genetic structure (SGS) can help identify intrinsic and extrinsic barriers to gene flow within metapopulations. For marine organisms such as coral reef fishes, identifying these barriers is critical to predicting evolutionary dynamics and demarcating evolutionarily significant units for conservation. In this study, we adopted an alternative hypothesis-testing framework to identify the patterns and predictors of SGS in the Caribbean reef fish *Elacatinus lori*. First, genetic structure was estimated using nuclear microsatellites and mitochondrial cytochrome b sequences. Next, clustering and network analyses were applied to visualize patterns of SGS. Finally, logistic regressions and linear mixed models were used to identify the predictors of SGS. Both sets of markers revealed low global structure: mitochondrial Φ_{ST} = 0.12, microsatellite F_{ST} = 0.0056. However, there was high variability among pairwise estimates, ranging from no differentiation between sites on contiguous reef (Φ_{ST} = 0) to strong differentiation between sites separated by ocean expanses ≥ 20 km (maximum Φ_{ST} = 0.65). Genetic clustering and statistical analyses provided additional support for the hypothesis that seascape discontinuity, represented by oceanic breaks between patches of reef habitat, is a key predictor of SGS in *E. lori*. Notably, the estimated patterns and predictors of SGS were consistent between both sets of markers. Combined with previous studies of dispersal in *E. lori*, these results suggest that the interaction between seascape continuity and the dispersal kernel plays an important role in determining genetic connectivity within metapopulations.

Keywords: connectivity, ecology, gene flow, marine, metapopulation, spatial genetic structure

Received 22 February 2014; revision received 25 April 2014; accepted 29 April 2014

Introduction

Current and historical patterns of gene flow influence the evolutionary dynamics of metapopulations and can result in the nonrandom distribution of alleles across space – spatial genetic structure (SGS) (Hanski & Gaggiotti 2004; Vekemans & Hardy 2004). Analysing the spatial distribution of alleles and gene lineages can therefore provide insight into historical and current genetic connectivity patterns (Avise 2000; Hellberg 2007) and can inform conservation planning through the identification of evolutionarily significant units or other management stocks (Schwartz *et al.* 2007; Cano *et al.* 2008). However, understanding the linkages between dispersal, gene flow and SGS is complicated. This is because real metapopulations are situated on complex landscapes, whose features interact with environmental variables and dispersal traits to influence genetic connectivity (Manel *et al.* 2003; Manel & Holderegger 2013). Therefore, investigating both the patterns and predictors of SGS is necessary to fully understand the microevolutionary consequences of gene flow.

In contrast to terrestrial systems, where the idea of spatial substructure within metapopulations has strongly influenced population genetic theory (Wright 1943), phylogeography (Avise 2000) and landscape genetics (Manel *et al.* 2003), early studies of genetic structure in marine systems operated under the hypothesis that

Correspondence: C. C. D'Aloia, Fax: 617 353 6340;
E-mail: cdaloia@bu.edu

marine metapopulations were relatively 'open' (Roberts 1997). Two assumptions about marine metapopulations gave rise to this hypothesis of 'openness': (i) that they would have high rates of effective dispersal and (ii) that they would have permeable extrinsic barriers to gene flow (Taylor & Hellberg 2006; Cano *et al.* 2008). Thus, pelagic larval dispersal over basin-wide scales was predicted to result in widespread panmixia, and many studies found limited evidence for population structure in the sea (see Shulman & Bermingham 1995; Purcell *et al.* 2006). However, within the last decade, there has been a paradigm shift regarding the scaling of connectivity in marine populations, whereby mean and modal dispersal distances for many species are now expected to fall within ten to one hundred kilometres (Jones *et al.* 2009). Evidence that has driven this shift comes from coupled biophysical models that predict restricted dispersal (Cowen *et al.* 2006), population genomic studies of structure that provide indirect evidence of restricted dispersal (Corander *et al.* 2013; Reitzel *et al.* 2013) and genetic parentage analyses that provide direct evidence of restricted dispersal (Buston *et al.* 2012; Almany *et al.* 2013; D'Aloia *et al.* 2013). These data indicate that strong extrinsic and intrinsic barriers to gene flow may exist in marine systems and that biologically meaningful genetic structure may be detectable.

To begin to identify barriers to dispersal and gene flow, alternative hypotheses of the predictors of SGS must be tested. To facilitate this, the landscape genetic analytical framework can be applied, with some of the hypotheses modified for the marine environment. At present, there are five major hypotheses, a null and four alternatives, for the drivers of neutral SGS in marine metapopulations. The null hypothesis posits that genetic structure is random in space. This idea has two plausible mechanisms: (i) larvae are passive propagules and larval cohorts mix completely in the plankton (Victor 1984), or (ii) marine dispersal is driven by 'sweepstakes reproduction' in which variance in individual reproductive success is high and stochastic, resulting in random directionality of gene flow over time (Hedgecock 1994). The first alternative hypothesis is that seascape discontinuity, or fragmentation of marine habitat patches, can restrict gene flow (Johnson & Black 1991). Additional seascape features, such as gradients in temperature or salinity, can also represent barriers (Rocha *et al.* 2007). This hypothesis is analogous to the isolation-by-barrier (IBB) hypothesis in landscape genetics. Second, there is a classic hypothesis from population genetics that genetic distance between populations increases with geographical distance (Wright 1943). This isolation-by-distance (IBD) hypothesis assumes that the probability of gene flow declines with Euclidian distance. Third, some studies have hypothesized that patterns of ocean currents, as a measure of physical connectivity, are a strong predictor of SGS (see Selkoe *et al.* 2006). Therefore, a more appropriate distance metric for marine organisms may be derived from oceanographic models – this hypothesis has been termed isolation-by-'derived oceanographic distance' (IBDOD) (White *et al.* 2010). Fourth, other studies have hypothesized that species-specific life history traits may restrict dispersal potential, thereby creating intrinsic barriers to gene flow (see Pelc *et al.* 2009). Considering these potential drivers of SGS, we can begin to disentangle their relative influence.

In this study, we use the neon goby *Elacatinus lori* as a tractable study organism for the application of this alternative hypothesis-testing framework to a marine species. *E. lori* is a suitable study organism for three reasons. First, it is an endemic to the Mesoamerican barrier reef system (MBRS), with the majority of its range constrained within Belize (Colin 2002; D'Aloia *et al.* 2011). This endemicity facilitates SGS measurement across a large proportion of the species' range. Second, *E. lori* has been shown to have restricted larval dispersal even in continuous habitat (D'Aloia *et al.* 2013). This restricted dispersal pattern suggests that *E. lori* life history traits may be intrinsic barriers to dispersal and indicates that gene flow may also be restricted between demes in the metapopulation (Fig. 1). Third, another species in the same genus, *Elacatinus evelynae*, was found to have remarkably high levels of genetic structure between island subpopulations separated by as little as 20 km (Taylor & Hellberg 2003). Combined, the two latter lines of evidence suggest that biologically meaningful levels of SGS may be detectable in *E. lori*, motivating a first investigation of the predictors of SGS in this marine metapopulation.

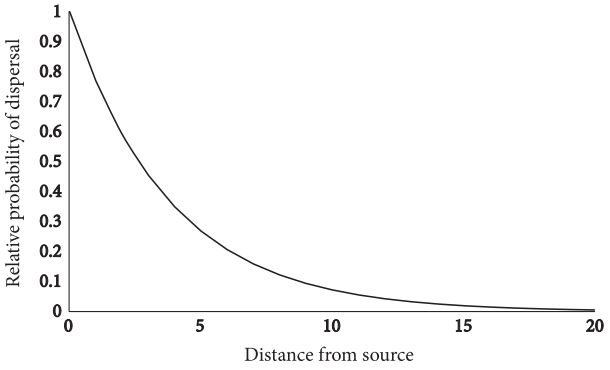

Fig. 1 First approximation of the *Elacatinus lori* dispersal kernel. The solid black line represents a first approximation of the kernel estimated from genetic parentage analysis. The kernel was measured up to 5 km from source; here, we extrapolate out to 20 km for comparison with spatial genetic structure data. Figure modified from D'Aloia *et al.* (2013).

Here, we characterize the patterns and predictors of SGS in *E. lori* across the Belizean portion of the MBRS. Using both nuclear and mitochondrial markers, we proceed in three steps. First, we conduct basic analyses of genetic structure to determine whether there is any evidence of significant differentiation between sampling sites. Second, we conduct qualitative clustering and network analyses to visualize the spatial pattern of structure. Third, we use logistic regression and mixed model statistical analyses to test three alternative hypotheses of the predictors of SGS: (H_0) pairwise genetic structure is randomly distributed in space; (H_1) pairwise genetic structure is associated with oceanic breaks between patches of reef habitat (IBB); (H_2) pairwise genetic structure is correlated with Euclidian distance (IBD). We did not test a third alternative hypothesis, H_3, which posits that ocean currents influence pairwise genetic structure (IBDOD), because this would require a high-resolution, hydrodynamic model that has not yet been developed for the region or a fourth alternative hypothesis, H_4, which posits that species traits influence SGS, because this would require the concurrent investigation of multiple species. The results of this study indicate that seascape continuity plays a predominant role in determining patterns of gene flow across this reef system. Further, the results provide support for the idea that geographical barriers to genetic connectivity can occur at a much smaller scale than has been assumed in reef fishes.

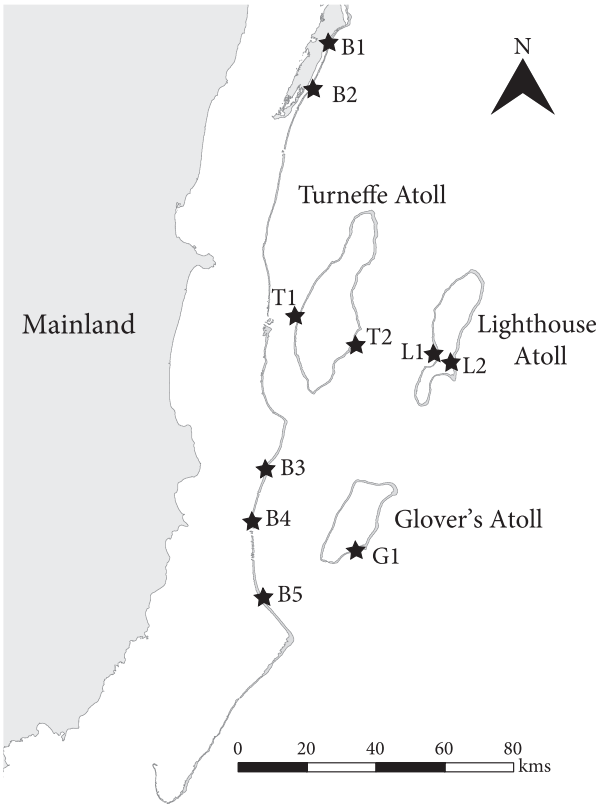

Fig. 2 Map of the Belizean barrier reef complex. Reef is represented by offshore grey lines. Sampling locations (n = 10) are represented by black stars. Barrier reef sites are labelled with a B, while atoll sampling sites are labelled with the first letter of their atoll name.

Methods

Tissue collection

To investigate population structure in *Elacatinus lori* across the Belize barrier reef complex (BBR), tissue samples were collected at ten sites along the barrier and atoll reefs (Fig. 2). There were a total of five barrier reef sites and five atoll reef sites. At each site, tissue samples from approximately 20–30 adult individuals were collected by divers using SCUBA. Each individual was removed from its host sponge using a slurp gun, and a small, nonlethal clip was cut from the caudal fin (D'Aloia *et al.* 2013). Individuals were returned to their sponges, and fin clips were stored in 95% ethanol upon dive completion.

Genotyping and sequencing

Genomic DNA was extracted using DNeasy Blood and Tissue Kits (Qiagen). Individuals were genotyped at 14 microsatellite loci according to the protocol detailed in D'Aloia *et al.* (2013). Diluted amplicons were screened on an ABI 3730 automated sequencer, and allele sizes were determined with GENEMAPPER v.4.0 (Life Technologies).

In addition to the nuclear microsatellite markers, a 1102-bp region of the cytochrome b mitochondrial gene (*cytb*) was amplified using *E. lori*-specific primers: Elori_cytbF (5′ - GGCCGCCCTACGAAAAACCC - 3′) and Elori_cytbR (5′ - TAGAGGGAAAAAGGCCAAGAAAA TAGAAA - 3′). Standard PCRs were run with the following reagents per reaction: 6.9 μL H_20, 1 μL 10× PCR buffer, 0.4 μL $MgCl_2$ (50 mM), 0.2 μL forward primer (10 μM), 0.2 μL reverse primer (10 μM), 0.2 μL dNTP (10 mM) and 0.1 μL Platinum Taq (Life Technologies). PCR thermal cycling followed a 'touchdown' protocol: six cycles of 95 °C for 40 s, 61–56 °C for 45 s (annealing temperature decreased 1 °C each cycle) and 72 °C for 1 min; 29 cycles of 95 °C for 40 s, 55 °C for 45 s and 72 for 1 min; and a final extension at 72 °C for 5 min. Amplicons were treated with an enzymatic cleanup to remove excess primers and dNTPs. The cleanup reagents added to each reaction included 4.3 μL H_20, 0.5 μL 10× PCR buffer, 0.1 μL shrimp alkaline phosphatase (GE Healthcare) and 0.1 μL exonuclease I (New

England Biolabs). Reactions were incubated at 37 °C for 45 min, followed by 90 °C for 15 min.

To sequence this region of *cytb*, we obtained forward and reverse sequences for every individual. For each sequencing reaction, 1 μL enzymatically cleaned amplicons was added to a mix of: 2.6 μL Sigma H_20, 0.5 μL BIGDYE TERMINATOR v. 3.1 Ready Reaction Mix (Life Technologies), 0.75 μL ABI 5× sequencing buffer and 0.15 μL primer (10 μM forward or reverse).

Sequencing reactions were cleaned using Agencourt CleanSeq beads (Beckman Coulter) and were sequenced on an ABI 3730 automated sequencer. Forward and reverse sequences were joined into contigs in CODONCODE ALIGNER v.4.0.4. All contigs were then compared with the ClustalW alignment algorithm (Thompson *et al.* 1994), and every contig was manually checked to remove false gaps and resolve ambiguous base calls. The region of *cytb* used for population structure analyses was restricted to 960 bp of high-quality sequence data.

Genetic summary statistics

Basic summary statistics were calculated for microsatellite and mitochondrial markers. For the microsatellites, the average number of alleles per locus (*k*), observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated for each sampling site in ARLEQUIN v.3.5.1.3 (Excoffier & Lischer 2010). Within each sampling site, more detailed statistics were also recorded, including tests of linkage disequilibrium (LD). Significance of LD was evaluated with permutation tests and a sequential Bonferroni correction to account for multiple pairwise comparisons. Deviations from Hardy–Weinberg equilibrium were also assessed using an exact test (chain length = 1001 000; burn-in = 10 000). Finally, the presence of null alleles was investigated in MICROCHECKER v.2.2.3 (van Oosterhout *et al.* 2004). Mitochondrial DNA polymorphism was summarized by characterizing the number of haplotypes, the number of segregating sites (S), the types of mutations (i.e. transitions, transversions or indels) and four estimates of theta.

Basic analyses of genetic structure

To test for the presence of genetic structure, we investigated global and pairwise structure between all sampling sites. First, we estimated Φ_{ST} for mitochondrial *cytb* sequences and F_{ST} for nuclear microsatellites, with significance evaluated by permutation tests (n = 10 100) (ARLEQUIN). For sequence data, Φ_{ST} can be a more informative estimator of pairwise differentiation than F_{ST} because it accounts for the genetic distance between haplotypes.

Recognizing the limitations of traditional metrics of differentiation (Hedrick 2005), we applied two additional approaches to the data. First, we estimated Hedrick's G'_{ST} from the microsatellite data, which can be a more appropriate measure of differentiation when heterozygosity is high. Hedrick's G'_{ST} corrects mathematically for the fact that F_{ST} declines as polymorphism increases. Second, we partitioned total genetic variation into within-site and among-site covariance components with an AMOVA for both sets of markers (ARLEQUIN).

Qualitative analyses of patterns of spatial genetic structure

After detecting evidence of pairwise population structure between sampling sites, we investigated qualitatively whether there was any spatial pattern of genetic structure. To look for patterns across the BBR, we used genetic clustering algorithms and haplotype network analyses. Using these approaches, the distribution of genetic clusters and/or haplotypes was then overlaid onto the reef locations where individuals were sampled to visualize SGS.

First, microsatellite data were used to estimate the number of genetic clusters present across the BBR (STRUCTURE v.2.3.4, Pritchard *et al.* 2000). We used an admixture model to account for historical and/or contemporary gene flow and the correlated allele frequencies model, which can better account for subtle signatures of structure. Clusters from k = 1 to k = 10 were tested. Each MCMC chain ran for 150 000 burn-in steps, followed by 100 000 additional steps, and 20 chains were run for each value of *k*. To choose among potential values of *k*, parameter estimates were pooled among runs, and alternative models were compared using the Evanno method (Evanno *et al.* 2005) as implemented in STRUCTURE HARVESTER v.0.6.93 (Earl & vonHoldt 2012). Upon selecting k, data across runs were optimally aligned in *clumpp* using the Greedy algorithm (input order = random; repeats = 1000) (Jakobsson & Rosenberg 2007). Finally, aligned data were visualized in *distruct* (Rosenberg 2004).

Second, a mitochondrial haplotype network was constructed using TCS v.1.21 with 95% parsimony (Clement *et al.* 2000). To visualize the spatial distribution of mitochondrial haplotypes, we used the network to group the 48 haplotypes into categories based on the genetic distance between them. Because the network revealed two predominant haplotypes ('1' and '2') that each had many closely related haplotypes, we binned the data into six categories: (i) 'haplotype 1'; (ii) 'haplotypes that were one mutation away from haplotype 1'; (iii) 'haplotypes that were two or more mutations away from haplotype 1'; (iv) 'haplotype 2'; (v) 'haplotypes that were one mutation away from haplotype 2'; and (vi) 'haplotypes that were two or more mutations away from hap-

lotype 2'. Next, we used pie charts to map the relative frequency of the six categories of haplotypes at each sampling site. We also conducted phylogenetic analyses, but trees could not be resolved due to low intraspecific sequence divergence (Appendix S1, Supporting information).

Quantitative analyses of the predictors of spatial genetic structure

To test alternative hypotheses of the predictors of SGS in the *E. lori* metapopulation, we used logistic regression and linear mixed model analyses. This two-step approach was necessary because zero inflation in the pairwise differentiation data sets precluded data transformation for normality, which is a fundamental assumption of linear models (Martin *et al.* 2005). First, we tested the predictors of a binary dependent variable that described whether pairwise differentiation was zero (0) or any positive value (1). Because there is quasi-complete separation in the data, whereby one or more covariates nearly perfectly predict some binary dependent variable, we applied a Firth penalized-likelihood logistic regression instead of a generalized linear mixed model (GLMM) with a logit link (R: *logistf* package). With standard logistic regression approaches, parameter estimates can approach infinity with separation, while the Firth bias correction uses a penalized estimation method that allows for consistent estimation of parameters, even in the presence of separation (Firth 1993). Second, after excluding zero values from the data set, we log-transformed the positive Φ_{ST} and F_{ST} values and applied linear mixed models (LMMs) to investigate whether these same variables predicted the magnitude of positive-only differentiation values (R: *lme4* and *AICcmodavg* packages). Here, we explicitly accounted for the nonindependence of pairwise differentiation estimates by adding two random effect variables that accounted for within-site variation.

Specifically, we built a group of nested models for each analysis with sets of predictor variables that corresponded with either the null hypothesis or the nonmutually exclusive alternative hypotheses (H_1 and H_2). The fit of alternative models was then determined using penalized-likelihood ratio tests and second-order Akaike information criterion (ΔAIC_c) for the Firth logistic regressions and LMMs, respectively. In this way, results could be interpreted as support or rejection of each hypothesis.

First, we tested the null hypothesis that SGS is randomly distributed in space (H_0). For the Firth logistic regressions, the null was tested in a frequentist statistical framework. For the LMMs, a specific null model was constructed with only two random effect variables – 'Site 1' and 'Site 2'.

Second, we tested the IBB hypothesis that pairwise genetic differentiation was correlated with oceanic breaks between reef patches (H_1). All pairs of sites were grouped by a dummy variable that described whether sites were separated by an oceanic break ≥ 20 km (1) or were situated on contiguous reefs (0). For example, a pair where both sites are on the barrier reef would be considered a contiguous pair (0), while a pair with one site on the barrier reef and one site on Glover's Atoll would be considered separated (1). This distance cut-off was chosen to quantitatively test the observation from genetic clustering algorithms that relatively isolated sites, separated by at least 20 km from other sites, were highly differentiated. However, shorter distance cut-offs were also used for robustness checks (Appendix S2, Supporting information). This model added a main predictor variable 'oceanic break'.

Third, we tested the IBD hypothesis that the observed genetic differentiation increases with geographic distance (H_2). A model was constructed that added 'Euclidian distance' as an additional predictor variable to test whether the addition of geographic distance improved model fit, while controlling for the influence of 'oceanic breaks'.

Results

Summary statistics

The microsatellite markers were highly polymorphic with the mean number of alleles per locus per site ranging from 13.36–17.07 (Appendix S1, Supporting information). Similarly, observed heterozygosity was high, ranging from 0.78 to 0.83. Within sites, few loci showed deviations from HWE (2–5 deviations per site), and MICROCHECKER analyses suggested that most of these deviations were attributable to null alleles (Appendix S1, Supporting information). Site T2 was the only site where there was any evidence for linkage disequilibrium between loci after a sequential Bonferonni correction; within T2, only 1 of 91 pairwise comparisons was significant (Appendix S1, Supporting information). Therefore, although some caution should be taken in interpreting results from T2, loci are treated as unlinked in further analyses.

There were 48 mitochondrial haplotypes identified among the 294 individuals sequenced, with a total of 50 polymorphic sites within a 960-bp region of cytb (5.2% of all sites). Across sampling locations, haplotypes exhibited more transitions than transversions, and there were no indels (Appendix S1, Supporting information).

Within sampling locations, the number of private substitution sites ranged from zero to five.

Basic analyses of genetic structure

There was evidence for significant genetic structure based on global and pairwise estimates for both microsatellite genotypes and mitochondrial sequences (Table 1). Global Φ_{ST} for mtDNA was 0.12, while global F_{ST} for microsatellites was 0.0056. Pairwise Φ_{ST} estimates ranged from zero ($\Phi_{ST} = 0$) to strongly differentiated ($\Phi_{ST} = 0.65$). For most pairs of sites, Φ_{ST} estimates based on mitochondrial haplotypes were substantively larger than F_{ST} estimates based on microsatellites, which can be attributed to the high levels of heterozygosity in the microsatellites. However, there was a significant correlation between these two genetic distance matrices (Mantel test: Spearman's rho = 0.64, permutations = 1000; $P = 0.007$). Sites located on the two most geographically isolated atolls, Lighthouse Atoll (L1) and Glover's Atoll (G1), exhibited elevated levels of pairwise differentiation. In particular, Lighthouse Atoll (L1) was the most genetically differentiated site (mtDNA range: $\Phi_{ST} = 0.11$–0.65).

Given the problems associated with using F_{ST} and its relatives as metrics of differentiation, we conducted additional analyses to test the robustness of our results. First, an alternative metric of differentiation based on microsatellites, G'_{ST}, revealed higher levels of pairwise differentiation between sampling locations relative to unstandardized F_{ST} estimates; however, they were still substantially smaller than estimates based on mtDNA (Appendix S1, Supporting information). Second, AMOVAs for both sets of markers indicated that much of the observed genetic variation is attributable to variation within sites as opposed to variation among sites (Table 2); however, this percentage varies among markers. For the microsatellite genotypes, virtually all of the variation (99.44%) is attributable to variation within sites, while less variation in the mitochondrial haplotypes (87.59%) is attributable to variation within sites. These results indicate that while caution should be used in inferring structure (or a lack thereof) from microsatellites, data from both sets of markers are consistent with the hypothesis that some pairs of sampling sites have had, or do have, restricted gene flow.

Qualitative analyses of patterns of spatial genetic structure

Clustering analysis in STRUCTURE revealed four distinct genetic clusters across the BBR based on the microsatellite genotypes. Each of the two most isolated atolls, Glover's and Lighthouse, is primarily composed of one unique genetic cluster ('pink' for Glover's Atoll and 'green' for Lighthouse Atoll), although there is some evidence for shared ancestral polymorphism and/or limited ongoing gene flow with other sites (Fig. 3a). Turneffe Atoll, particularly site T1, appears to be a mixing site for multiple genetic clusters ('pink', 'green', 'yellow' and 'blue'); this could be explained by its proximity to the barrier reef and its central location relative to all other major reef regions (Fig. 2). In contrast, sites along the barrier reef are predominantly characterized by two different genetic clusters ('yellow' and 'blue') (Fig. 3a). There is some evidence for a cline in neutral markers along the barrier reef: individuals from northern sites (B1, B2) have a large fraction of their genotype assigned to the 'yellow' genetic cluster, while individuals from the southern sites (B3, B4, B5) have a much larger fraction assigned to the 'blue' genetic cluster. These results are interesting because STRUCTURE does not use a priori knowledge of sampling site coordinates. Taken together, these data suggest that large oceanic breaks (i.e. the breaks separating Glover's and Lighthouse Atolls from other reef patches) may represent strong extrinsic barriers to gene flow for *E. lori*, which is consistent with predictions from its restricted dispersal

Table 1 Pairwise population structure. Lower triangle of the matrix shows pairwise Φ_{ST} using *cytb* mtDNA sequences. The upper triangle of the matrix shows pairwise F_{ST} using 13 microsatellite loci (1419tri was excluded due to null alleles). Negative values were adjusted to zero. Significant values are bolded, with significance assessed by a permutation test at $\alpha = 0.05$ (10 100 permutations)

	B1	B2	B3	B4	B5	T1	T2	G1	L1	L2
B1	—	0.0007	0	0.0001	**0.0055**	0	0	**0.0083**	**0.0085**	**0.0077**
B2	0.0762	—	0.0033	0.0029	0.0047	0.0050	0.0037	**0.0136**	**0.0050**	**0.0117**
B3	0	0.0677	—	0	0.0030	0	0.0023	**0.0121**	**0.0138**	0.0053
B4	0	0.0204	0	—	0.0024	0.0009	0.0025	**0.0102**	**0.0088**	**0.0075**
B5	0	0.0053	0	0	—	0.0048	0.0048	**0.0089**	**0.0048**	0.0040
T1	0.0085	0	0	0	0	—	0	**0.0052**	**0.0107**	**0.0058**
T2	0	0.0165	0	0	0	0	—	**0.0101**	**0.0120**	**0.0111**
G1	0.0296	**0.2598**	0.0549	**0.1033**	**0.1250**	**0.1427**	**0.1010**	—	**0.0132**	**0.0148**
L1	**0.3248**	**0.6453**	**0.3892**	**0.4309**	**0.4609**	**0.4729**	**0.4078**	**0.1446**	—	0.0056
L2	0.0678	**0.3329**	**0.1085**	**0.1584**	**0.1836**	**0.2022**	**0.1510**	0	**0.1090**	—

Table 2 Analysis of molecular variance (AMOVA). AMOVAS partition total variance into covariance components for within and among populations. Results are presented based on (a) mitochondrial cytb haplotypes and (b) 13 nuclear microsatellites. Significance tests by permutation reveal $P < 0.001$ for both sets of markers (1023 permutations)

(a) cytb mtDNA					(b) Nuclear microsatellites				
Source of variation	d.f.	Sum of squares	Variance components	% variation	Source of variation	d.f.	Sum of squares	Variance components	% variation
Among sites	9	66.565	0.20351 Va	12.41%	Among sites	9	66.941	0.03129 Va	0.56%
Within sites	284	407.772	1.43582 Vb	87.59%	Within sites	586	3269.104	5.57868 Vb	99.44%
Total	293	474.337	1.63933	100%	Total	595	3336.045	5.60997	100%

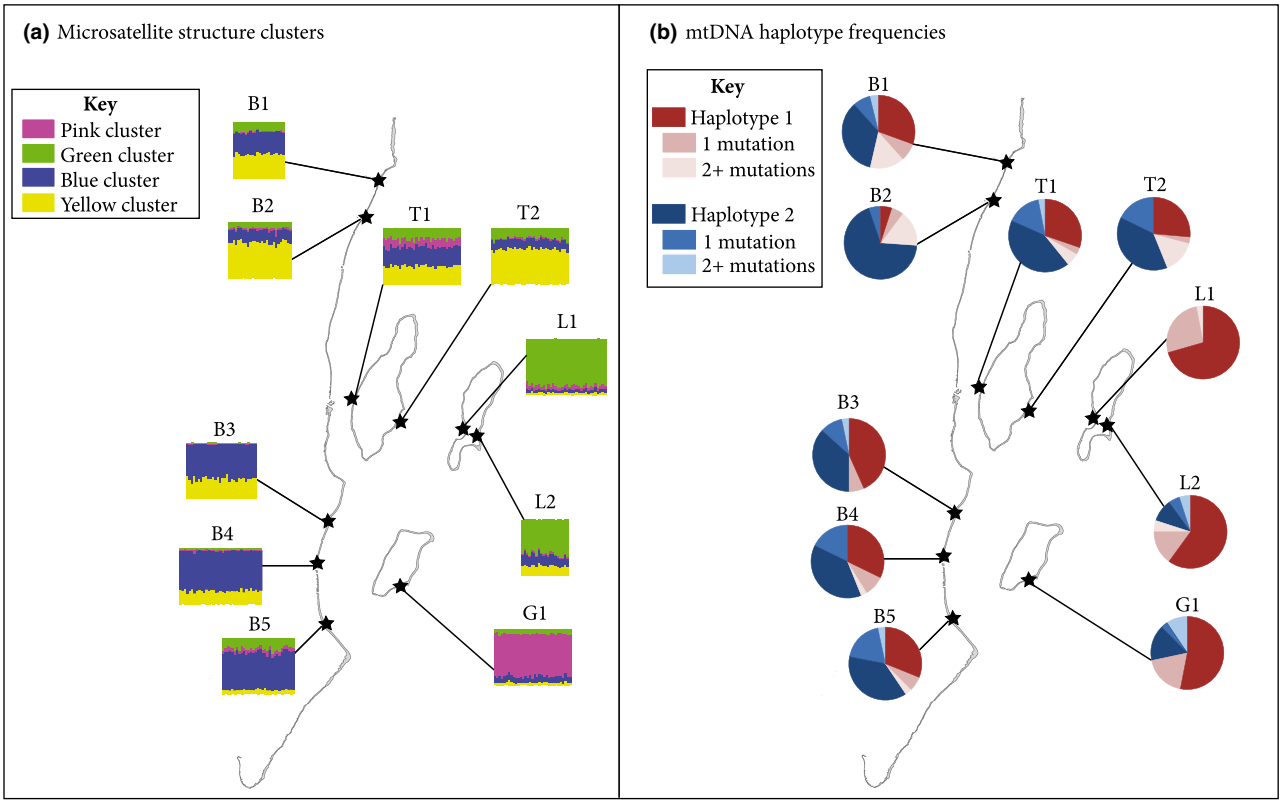

Fig. 3 Qualitative analyses of spatial genetic structure. (a) Distribution of $k = 4$ genetic clusters across all ten sampling sites, based on STRUCTURE analysis of microsatellites. These data show distinct clusters on Glover's Atoll (predominantly pink) and Lighthouse Atoll (predominantly green). The third atoll – Turneffe – contains a mix of all clusters, particularly at site T1. Along the barrier reef, there are two predominant clusters (yellow and blue). Vertical bars within each site represent individuals. (b) Relative frequencies of mitochondrial haplotypes across sampling sites. The 48 *cytb* haplotypes were grouped into six categories based on the haplotype network (Appendix S1, Supporting information). Beginning with the two predominant haplotypes (1 and 2), we then defined haplotypes as either 1 or 2+ mutations away from each of the predominant haplotypes. Haplotype 1 and its relatives are depicted in shades of red, while haplotype 2 and its relatives are depicted in shades of blue.

curve (Fig. 1; D'Aloia *et al.* 2013). In contrast, some form of lattice dispersal (whereby individuals move to neighboring patches in a network) may occur each generation along the relatively continuous barrier reef.

The mitochondrial haplotype network also supports the hypothesis that isolated atolls are genetically differentiated from other sites. The haplotype network reveals that there are two predominant haplotypes, separated by only five point mutations (Appendix S1, Supporting information). There is a high frequency of rare haplotypes, many of which occur at only one locality and are separated from one of the predominant haplotypes by only one or two mutations. Mapping the relative frequencies of the predominant haplotypes and

their relatives demonstrates a clear spatial pattern in haplotype distribution: while sites on the barrier reef and Turneffe Atoll are characterized by a mixture of all haplotype categories, sites on the remote Lighthouse and Glover's Atolls are predominantly characterized by haplotype 2 and its relatives (Fig. 3b). Most notably, at L1, the site on Lighthouse Atoll, which appears to be the most differentiated from all other sites, haplotype 1 and its relatives are completely absent.

Quantitative analyses of the predictors of spatial genetic structure

Logistic regression and mixed model analyses were used to test alternative hypotheses of the predictors of SGS, and the results support the hypothesis that SGS is correlated with oceanic breaks between patches of reef habitat (H_1). Here, we focus on models predicting Φ_{ST} (for mtDNA) and F_{ST} (for nuclear microsatellites).

Firth's penalized-likelihood logistic regression revealed that sites separated by an oceanic break ≥ 20 km are 31 times more likely to have a Φ_{ST} estimate greater than zero, as compared to sites that are not separated by a break [coefficient = 3.43 (odds ratio = 31.00); SE = 0.97; $P < 0.001$]. There was no significant improvement over this 'oceanic break' model when 'Euclidian distance' was included as a predictor ($\chi^2 = 0.77$, d.f. = 1, $P = 0.38$). These results were robust to changing the structure metric to microsatellite-based F_{ST}. Here, the best-fit model revealed that sites separated by a break were over 18 times more likely to have a nonzero value of F_{ST} relative to sites that are not separated by a break [coefficient = 2.92 (odds ratio = 18.52); SE = 1.53; $P = 0.007$] and that adding Euclidian distance as a predictor did not improve model fit ($\chi^2 = 1.12$, d.f. = 1, $P = 0.29$). Taken together, these results reject the ideas that SGS is random (H_0) or that SGS follows an isolation-by-distance pattern (H_2), but provide support for H_1 that posits that SGS is associated with oceanic breaks.

Building upon these findings, linear mixed models revealed that oceanic breaks ≥ 20 km were also significantly associated with the magnitude of nonzero structure estimates between sites. The presence of a break was associated with a 93% increase in pairwise Φ_{ST} (Table 3a) and a 56% increase in pairwise F_{ST} (Table 3b), relative to pairs of sites without a break. For Φ_{ST} models, the 'ocean break' model (H_1; $\Delta AIC_c = 0.00$) was a better fit than a null model of only random effects (H_0; $\Delta AIC_c = 2.51$), and an alternative model that also included Euclidian distance (H_2; $\Delta AIC_c = 2.69$). A comparison of F_{ST} models revealed the same pattern: the 'ocean break' model (H_1; $\Delta AIC_c = 0.00$) was a better fit than a null model of only random effects (H_0; $\Delta AIC_c = 6.00$), and an alternative model that also included Euclidian distance (H_2; $\Delta AIC_c = 2.75$). Taken together, these results show strong statistical support for H_1 across both genetic markers.

Importantly, these results were not robust to shortening the distance cut-off for the definition of an 'oceanic break' (Appendix S2, Supporting information). This result is congruent with the qualitative clustering analyses. Together, they demonstrate that only substantively 'large' oceanic breaks (≥20 km) are associated with SGS. In sum, the results support the hypothesis that large oceanic breaks are significant predictors of SGS for *E. lori* (H_1) and, by inference, may be barriers to larval dispersal and gene flow for *E. lori*.

Discussion

Investigating the patterns and predictors of spatial genetic structure (SGS) concurrently is essential to

Table 3 Linear mixed model output. The dependent variable is nonzero values of pairwise Φ_{ST} (a) or F_{ST} (b), logged for normality. We report parameter estimates, standard errors and t values for the fixed effects in the best-fit model from the group of nested models. We also present standard deviations for the random effect intercepts

(a) Φ_{ST} (mtDNA)				(b) F_{ST} (microsatellites)			
Fixed effects				*Fixed effects*			
Parameter	**Estimate**	**SE**	***t* value**	**Parameter**	**Estimate**	**SE**	***t* value**
(Intercept)	−1.69	0.20	−8.28	(Intercept)	−2.62	0.08	−33.46
Oceanic break > 20 km	0.93	0.34	2.74	Oceanic break > 20 km	0.56	0.10	5.45
Random effects				*Random effects*			
Group	**SD**			**Group**	**SD**		
Site 1 (intercept)	0.22			Site 1 (intercept)	0.00		
Site 2 (intercept)	0.48			Site 2 (intercept)	0.00		

understanding genetic connectivity within metapopulations. In turn, understanding connectivity is important because it influences evolutionary dynamics and the delineation of conservation units. The emerging field of seascape genetics, which applies the landscape genetic analytical framework to marine organisms, is a powerful approach to identifying the environmental and biophysical drivers of SGS (Selkoe *et al.* 2008; Manel & Holderegger 2013). In this study, we conducted a preliminary seascape genetic analysis of the patterns and predictors of SGS in a reef fish, *E. lori*. Estimates of global structure were low across 10 sites on the Belizean barrier reef, but there was a wide range in pairwise divergence. Qualitative clustering analyses indicated that geographically isolated sites, separated from other reef sites by large oceanic breaks, were the most genetically differentiated sites. This finding was supported by statistical analyses that identified oceanic breaks between reef habitat patches as a significant predictor of SGS in *E. lori*. The alternative potential predictor – Euclidian distance – was not found to significantly improve model fit. These results suggest that discontinuity of the seascape may play an important role in creating barriers to gene flow in this reef fish.

These results are consistent with the groundbreaking population genetic research on the Caribbean genus *Elacatinus* (Taylor & Hellberg 2003, 2006). These studies revealed remarkably high levels of population structure among island populations in three other *Elacatinus* spp. (maximum $\Phi_{ST} > 0.7$) and inferred restricted dispersal as a possible mechanism driving these high estimates of structure. However, while F_{ST} is a useful metric of genetic structure, inferring a causal relationship between F_{ST} and dispersal is generally problematic because (i) F_{ST} can be influenced by multiple processes, including selection, inbreeding and drift (in addition to spatial subdivision and dispersal), and (ii) the assumptions of theoretical models relating pairwise F_{ST} to dispersal are nearly always violated in natural populations (Whitlock & McCauley 1999).

Given the limitations associated with inferring dispersal from measures of F_{ST}, a complete understanding of the linkages between the two requires measures of both a dispersal kernel and genetic structure in the same system. Because of the challenges involved, it is only relatively recently that a few studies have begun to directly quantify marine larval dispersal kernels via genetic parentage analysis (Buston *et al.* 2012; Almany *et al.* 2013). *E. lori* is one of the few marine species in which a dispersal kernel has been directly estimated (D'Aloia *et al.* 2013) and the only species for which we now have both the kernel and SGS data (presented here). The first approximation of the *E. lori* dispersal kernel revealed a rapid exponential decline in the probability of dispersal with respect to distance from source (Fig. 1; D'Aloia *et al.* 2013). Thus, with these two data sets, we can begin to integrate marine dispersal and SGS data for the first time in a reef fish to explicitly test their relationship, and to generate new insights about population connectivity.

Because *E. lori* is distributed on both offshore atoll reefs and along 250+ km of the relatively continuous barrier reef in Belize, patterns of dispersal and SGS can be compared in two types of seascapes. First, in concordance with SGS patterns in other *Elacatinus* spp., 20-km distances across open ocean are associated with high differentiation. This isolation-by-barrier (IBB) pattern may be explained by the dispersal kernel: the relative probability of dispersal tends to zero by 20 km from source, suggesting that a larva has a very low probability of successfully traversing such a wide ocean expanse (Fig. 1). However, 20-km distances along the continuous barrier reef are not associated with pairwise differentiation. This lack of an isolation-by-distance (IBD) pattern may be explained by the high probability of larvae connecting adjacent populations on the barrier reef over multiple generations (i.e. through stepping-stone dispersal). Together, these results show that the interaction between a species' dispersal kernel and habitat continuity can explain variation in SGS across a heterogeneous seascape.

Looking beyond *Elacatinus*, our results are consistent with previous research that has found a relationship between seascape continuity and genetic structure in other taxa. Since the first empirical study explicitly linked marine habitat continuity to SGS in a marine gastropod with direct development (Johnson & Black 1991), similar patterns have emerged in species with a dispersive propagule phase. Organisms as diverse as kelp with dispersive spores (Billot *et al.* 2003; Alberto *et al.* 2010) and marine fish with a pelagic larval phase (Johnson *et al.* 1994; Riginos & Nachman 2001) have also been shown to exhibit elevated pairwise differentiation when habitat patches are isolated. Thus, there is growing evidence that habitat continuity is a key predictor of genetic connectivity, which has important implications for marine conservation planning.

Interestingly, our findings deviate from previous research in regard to the IBD hypothesis. The subset of prior studies that also adopted a multivariate approach to investigating the predictors of SGS reported statistical support for the *combined* effects of habitat continuity (IBB) and geographical distance (IBD) (700+ km, Riginos & Nachman 2001; 700+ km, Billot *et al.* 2003; 70+ km, Alberto *et al.* 2010). In contrast, there was no statistical support for Euclidian distance as an additional predictor of SGS in *E. lori* (160+ km). One potential explanation for this disparity is that the IBD pattern observed

at large spatial scales in other studies may actually represent the accumulation of IBB effects at smaller spatial scales; however, testing this hypothesis will require high-resolution sampling and habitat mapping and should account for each species' dispersal potential. An alternative explanation is that there may be a subtle IBD pattern in *E. lori* that was not detected by our sampling scheme: the qualitative genetic clustering analysis for *E. lori* indicates, but does not conclusively show, a potential cline in neutral microsatellite markers along the continuous barrier reef (Fig. 3a). To more rigorously test the IBD hypothesis in this metapopulation, more intensive sampling along the reef will be required.

Notably, this study also demonstrates that while the same overall patterns of SGS may be detected by different genetic markers, the estimated magnitude of structure (and concomitant biological interpretation) may vary. Historically, population genetic studies of marine organisms have tended to use microsatellite markers, which have often revealed weak, but statistically significant levels of SGS. One issue with using microsatellites exclusively is that high levels of heterozygosity can lead to among-individual variation masking among-site variation (Hellberg 2007). Indeed, AMOVA results for *E. lori* reveal that nearly all of the microsatellite genetic variation is partitioned into variation among individuals, and our estimates of structure based on microsatellites were an order of magnitude lower than estimates based on mitochondrial DNA (Table 1). Despite this difference in magnitude, the results of the structure analyses, the clustering analyses and the statistical modelling were all consistent between microsatellite and mtDNA data. These congruent results between two sets of markers provide robust support for the hypothesis that large oceanic breaks are significant predictors of SGS for *E. lori*. They also suggest that the weak but significant structure detected by many microsatellite-based analyses of SGS in the sea (e.g. Purcell *et al.* 2006) may correlate with higher degrees of cryptic structure that could be revealed through alternative markers and/or expanded genome coverage (e.g. Corander *et al.* 2013; Reitzel *et al.* 2013).

Ultimately, a comprehensive understanding of the patterns and drivers of SGS in complex marine metapopulations can be achieved through a seascape genomics analytical framework that tests alternative hypotheses. To fully develop this framework, the integration of three additional types of data will be critical. First, environmental gradients, such as temperature and salinity, must also be considered as potential extrinsic barriers to gene flow (IBB). Second, bio-physical oceanographic models will enable the development of alternative metrics of distance based on ocean flow fields (IBDOD) (Cowen *et al.* 2006; White *et al.* 2010). Third, empirical estimates of dispersal kernels will capture species-specific dispersal potential and could facilitate the integration of intrinsic barriers to gene flow (Buston *et al.* 2012; Almany *et al.* 2013; D'Aloia *et al.* 2013). Considered together, these data will allow a comprehensive test of all the alternative hypotheses for SGS, enabling researchers to disentangle the relative effects of environmental heterogeneity, dynamic ocean currents, species-specific dispersal capabilities and their interactions on patterns of SGS.

Acknowledgements

We thank John Majoris, Alissa Rickborn and Kevin David for assistance in the field as well as the staff at Carrie Bow Caye and Calabash Caye Field Stations. We thank three anonymous reviewers for helpful comments on an earlier version of the manuscript. CCD was supported by a NSF GRF (Grant No. DGE-1247312), and the project was funded by a start-up award to PMB from the Trustees of Boston University. Research was approved by Belize Fisheries and the Boston University IACUC.

References

Alberto F, Raimondi PT, Reed DC *et al.* (2010) Habitat continuity and geographic distance predict population genetic differentiation in giant kelp. *Ecology*, **91**, 49–56.

Almany GR, Hamilton RJ, Bode M *et al.* (2013) Dispersal of grouper larvae drives local resources sharing in a coral reef fishery. *Current Biology*, **23**, 626–630.

Avise JC (2000) *Phylogeography. The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Billot C, Engel CR, Rousvoal S, Kloareg B, Valero M (2003) Current patterns, habitat discontinuities and population genetic structure: the case of the kelp *Laminaria digitata* in the English Channel. *Marine Ecology Progress Series*, **253**, 111–121.

Buston PM, Jones GP, Planes S, Thorrold SR (2012) Probability of successful larval dispersal declines fivefold over 1 km in a coral reef fish. *Proceedings of the Royal Society of London, Series B*, **279**, 1883–1888.

Cano JM, Shikano T, Kuparinen A, Merila J (2008) Genetic differentiation, effective population size and gene flow in marine fishes: implications for stock management. *Journal of Integrative Field Biology*, **5**, 1–10.

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Colin PL (2002) A new species of sponge-dwelling *Elacatinus* (*Pisces*: *Gobiidae*) from the western Caribbean. *Zootaxa*, **106**, 1–7.

Corander J, Majander KK, Cheng L, Merila J (2013) High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Molecular Ecology*, **22**, 2931–2940.

Cowen RK, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. *Science*, **311**, 522–527.

D'Aloia CC, Majoris JE, Buston PM (2011) Predictors of the distribution and abundance of a tube sponge and its resident goby. *Coral Reefs*, **30**, 777–786.

D'Aloia CC, Bogdanowicz SM, Majoris JE, Harrison RG, Buston PM (2013) Self-recruitment in a Caribbean reef fish: a

method for approximating dispersal kernels accounting for seascape. *Molecular Ecology*, **22**, 2563–2572.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

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, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Firth D (1993) Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.

Hanski IA, Gaggiotti OE (eds) (2004) *Ecology, Genetics and Evolution of Metapopulations*. Elsevier Academic Press, London.

Hedgecock D (1994) Temporal and spatial genetic structure of marine animal populations in the California Current. *California Cooperative Oceanic Fisheries Investigations Reports*, **35**, 73–81.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hellberg ME (2007) Footprints on water: the genetic wake of dispersal among reefs. *Coral Reefs*, **26**, 463–473.

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.

Johnson MS, Black R (1991) Genetic subdivision of the intertidal snail *Bembicium vittatum* (*Gastropoda*: *Littorinidae*) varies with habitat in the Houtman Abrolhos Islands, Western Australia. *Heredity*, **67**, 205–213.

Johnson MS, Watts RJ, Black R (1994) High levels of genetic subdivision in peripherally isolated populations of the atherinid fish *Craterocephalus capreoli* in the Houtman Abrolhos Islands, Western Australia. *Marine Biology*, **119**, 179–184.

Jones GP, Almany GR, Russ GR *et al.* (2009) Larval retention and connectivity among populations of corals and reef fishes: history, advances and challenges. *Coral Reefs*, **28**, 307–325.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology and Evolution*, **28**, 614–621.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.

Martin TG, Wintle BA, Rhodes JR *et al.* (2005) Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8**, 1235–1246.

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.

Pelc RA, Warner RR, Gaines SD (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography*, **36**, 1881–1890.

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

Purcell JFH, Cowen RK, Hughes CR, Williams DA (2006) Weak genetic structure indicates strong dispersal limits: a tale of two coral reef fish. *Proceedings of the Royal Society of London, Series B*, **273**, 1483–1490.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, **22**, 2953–2970.

Riginos C, Nachman MW (2001) Population subdivision in marine environments: the contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, *Axoclinus nigricaudus*. *Molecular Ecology*, **10**, 1439–1453.

Roberts CM (1997) Connectivity and management of Caribbean coral reefs. *Science*, **278**, 1454–1457.

Rocha LA, Craig MT, Bowen BW (2007) Phylogeography and the conservation of coral reef fishes. *Coral Reefs*, **26**, 501–512.

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

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution*, **22**, 25–33.

Selkoe KA, Gaines SD, Caselle JE, Warner RR (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology*, **87**, 3082–3094.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, **9**, 363–377.

Shulman MJ, Bermingham E (1995) Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution*, **49**, 897–910.

Taylor MS, Hellberg ME (2003) genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science*, **299**, 107–109.

Taylor MS, Hellberg ME (2006) Comparative phylogeography in a genus of coral reef fishes: biogeographic and genetic concordance in the Caribbean. *Molecular Ecology*, **15**, 695–707.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **2**, 4673–4680.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Victor BC (1984) Coral reef fish larvae: patch size estimation and mixing in the plankton. *Limnology and Oceanography*, **29**, 1116–1119.

White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ (2010) Ocean currents help explain population genetic structure. *Proceedings of the Royal Society of London, Series B*, **277**, 1685–1694.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: FST≠ 1/(4Nm+ 1). *Heredity*, **82**, 117–125.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

C.C.D. and P.M.B. planned the study. C.C.D. conducted field work with assistance from P.M.B. and others. C.C.D. conducted laboratory work with supervision by S.M.B. and R.G.H. C.C.D. analysed the data and wrote the paper, with all authors contributing to revisions.

Data accessibility

Microsatellite genotypes, mtDNA haplotype frequencies, mtDNA sequence alignment, sampling coordinates, and complete R code with associated data files are available in DRYAD (doi:10.5061/dryad.td28r). Individual sequences are available in GenBank (KF928971–KF929020).

Supporting information

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

Appendix S1 Summary statistics, alternative structure metrics, haplotype network, and phylogenetic analyses.

Appendix S2 Robustness checks for statistical analyses.